A proteome-wide screen utilizing second generation sequencing for the identification of lysine and arginine methyltransferase protein interactions

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m

(Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

Diplom-Biologin Mareike Weimann

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Dr. Andreas Herrmann

Gutachter: 1. Prof. Dr. Christian Spahn
2. Prof. Dr. Erich Wanker
3. Prof. Dr. Hans Lehrach

Tag der mündlichen Prüfung: 01. August 2012

Zusammenfassung

Proteinmethylierung spielt eine immer größere Rolle in der Regulierung zellulärer Prozesse. Die Entwicklung effizienter proteomweiter Methoden zur Detektion von Methylierung auf Proteinen ist limitiert und technisch schwierig. In dieser Arbeit haben wir einen neuen Hefe-Zwei-Hybrid-Ansatz (Y2H) entwickelt, der Proteine, die miteinander wechselwirken, mit Hilfe von Sequenzierungen der zweiten Generation identifiziert (Y2H-Seq). Der neue Y2H-Seq-Ansatz wurde systematisch mit dem Y2H-Seq-Ansatz verglichen. Dafür wurde ein Bait-Set von 8 Protein-Arginin-Methyltransferasen, 17 Protein-Lysin-Methyltransferasen und 10 Demethylasen gegen 14,268 Prey-Proteine getestet. Der Y2H-Seq-Ansatz ist weniger arbeitsintensiv, hat eine höhere Sensitivität als der Standard Y2H-Matrix-Ansatz und ist deshalb besonders geeignet, um schwache Interaktionen zwischen Substraten und Protein-Methyltransferasen zu detektieren. Insgesamt wurden 523 Wechselwirkungen zwischen 22 Bait-Proteinen und 324 Prey-Proteinen etabliert, darunter 11 bekannte Methyltransferasen-Substrate. Netzwerkanalysen zeigen, dass Methyltransferasen bevorzugt mit Transkriptionsregulatoren, DNA- und RNA-Bindeproteinen wechselwirken. Diese Daten repräsentieren das erste proteomweite Wechselwirkungsnetzwerk über Protein-Methyltransferasen und dienen als Ressource für neue potentielle Methylierungssubstrate. In einem *in vitro* Methylierungsassay wurden exemplarisch mit Hilfe massenspektrometrischer Analysen die methylierten Aminosäurereste einiger Kandidatenproteine bestimmt. Von neun getesteten Proteinen waren sieben methyliert, zu denen gehören SPIN2B, DNAJA3, QKI, SAMD3, OFCC1, SYNCRIP und WDR42A. Wahrscheinlich sind viele Methylierungssubstrate im Netzwerk vorhanden. Das vorgestellte Protein-Protein-Wechselwirkungsnetzwerk zeigt, dass Proteinmethylierung sehr unterschiedliche zelluläre Prozesse beeinflusst und ermöglicht die Aufstellung neuer Hypothesen über die Regulierung Molekularer Mechanismen durch Methylierung.

Schlagwörter: Proteinmethylierung, Protein-Arginin-Methyltransferase (PRMT), Protein-Lysin-Methyltransferase (PKMT), Demethylase, Protein-Protein-Wechselwirkung (PPI), Hefe-Zwei-Hybrid-System (Y2H), Sequenzierung der zweiten Generation, Protein-Wechselwirkungsnetzwerk

Abstract

Protein methylation on arginine and lysine residues is a largely unexplored posttranslational modification which regulates diverse cellular processes. The development of efficient proteome-wide approaches for detecting protein methylation is limited and technically challenging. We developed a novel workload reduced yeast-two hybrid (Y2H) approach to detect protein-protein interactions utilizing second generation sequencing. The novel Y2H-seq approach was systematically evaluated against our state of the art Y2H-matrix screening approach and used to screen 8 protein arginine methyltransferases, 17 protein lysine methyltransferases and 10 demethylases against a set of 14,268 proteins. Comparison of the two approaches revealed a higher sensitivity of the new Y2H-seq approach. The increased sampling rate of the Y2H-seq approach is advantageous when assaying transient interactions between substrates and methyltransferases. Overall 523 interactions between 22 bait proteins and 324 prey proteins were identified including 11 proteins known to be methylated. Network analysis revealed enrichment of transcription regulator activity, DNA- and RNA-binding function of proteins interacting with protein methyltransferases. The dataset represents the first proteome-wide interaction network of enzymes involved in methylation and provides a comprehensively annotated resource of potential new methylation substrates. An *in vitro* methylation assay coupled to mass spectrometry revealed amino acid methylation of candidate proteins. Seven of nine proteins tested were methylated including SPIN2B, DNAJA3, QKI, SAMD3, OFCC1, SYNCRIP and WDR42A indicating that the interaction network is likely to contain many putative methyltransferase substrate pairs. The presented protein-protein interaction network demonstrates that protein methylation is involved in diverse cellular processes and can inform hypothesis driven investigation into molecular mechanisms regulated through methylation.

Keywords: Protein methylation, protein arginine methyltransferase (PRMT), protein lysine methyltransferase (PKMT), demethylase, protein-protein interaction (PPI), yeast-two hybrid system (Y2H), second/next generation sequencing, PPI network

Abbreviations

AD	Activation domain
aDMA	Asymmetrically dimethylated arginine
AdoMet	S-adenosylmethionine
AdoHcy	S-adenosylhomocysteine
Adox	Periodate-oxidized adenosine
ALS6	Amyotrophic lateral sclerosis
AP-MS	Affinity purification coupled to mass spectrometry
ATP	Adenosintriphosphate
ATPase	Adenosintriphosphatasen
bp	Base pair
CoA	Coenzyme A
co-IP	Co-immunoprecipitation
CRL	Cullin-RING finger ligase
CTD	Carboxy-terminal domain
DBD	DNA-binding domain
DMEM	Dulbecco's modified eagle's medium
DNA	Deoxyribonucleic acid
dNTPs	Desoxyribonukleosidtriphosphate
DPBS	Dulbecco's phosphate buffered solution
ECL	Enhanced chemiluminescence
EGF	Epidermal growth factor
ES	Embryonic stem
ER	Estrogen receptor
FBS	Fetal bovine serum
FL	Full length
FNR	False negative rate
FPR	False positive rate
H	Histone
HD1	Homology domain 1
HD2	Homology domain 2
HEK	Human embryonic kidney
His	Histidine
HIV	Human immunodeficiency virus
hnRNPs	Heterogeneous nuclear ribonucleoproteins
HRP	Horseradish peroxidase
HTP	High-throughput

Ig	Immunoglobulin
IP	Immunoprecipitation
JmjC	Jumonji C
K	Lysine
KH	K homology
KRAB	Krueppel-associated box
LB	Lysogeny broth
LC-MS/MS	Liquid chromatographie-massenspektometrie/massenspektometrie
MMA	Monomethylated arginine
MPIMG	Max Planck Institute for Molecular Genetics
mRNA	Messenger RNA
MTP	Microtiterplate
Myr	Myristoylation morif
NCBI	National Centre for Biotechnological information
NF?B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NHR	Nuclear hormone receptor
NITE	National Institute of Technology and Evaluation
NLS	Nuclear localisation signal
OD600	Optical density at 600 nm
ORF	Open reading frames
PA	Protein A
PCR	Polymerase chain reaction
PDeM	Protein demethylase
PHD	Plant homeo domain
PKMT	Protein lysine methyltransferase
PMT	Protein methyltransferase
PPIs	Protein-protein interactions
PRMT	Protein arginine methyltransferase
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
R	Arginine
RG	Arginine-glycine
RING	Really interesting new gene
RLU	Relative luciferase units
RNA	Ribonucleic acid
RRM	RNA recognition motifs
SAM	S-adenosylmethionine
sDMA	Symmetrically dimethylated arginine

SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
seq	Sequencing
SET	Su(var), Enhancer of zest, trithorax
SH3	Src homology 3
snRNP	Small nuclear Ribonucleoprotein particles
SOCS	Suppressor of cytokines signalling
SR	Serine arginine
SRPY	SPla and the RYanodine Receptor
STAR	Transduction and activation of RNA
TBE	Tris Borate EDTA
TBS	Tris-buffered saline
TBST	TBS supplemented with Tween 20
TPR	Tetratricopeptide repeats
TSS	Transformation and storage solution
ub	Ubiquitin
UV	Ultra violet
Y2H	Yeast-two hybrid
ZnF	Zink finger

1 Introduction

1.1 Protein methylation

Protein methylation occurs predominantly on arginine, lysine and histidine residues and is catalyzed by S-adenosylmethionine (SAM, AdoMet) dependent enzymes that donate a methyl group to the side-chain nitrogen atom. Methylation reduces the charge and changes the structure of the side chain. Therefore it can alter function by increasing hybdrophobicity and may disrupt intra- and intermolecular hydrogen-bond interactions or provide a novel interface for proteins that bind preferentially to methylated proteins (Gary and Clarke, 1998; McBride and Silver, 2001). Enzymes catalyzing this reaction are called methyltransferases. Around 1 % of the human genome encode for methyltransferases with most of them presumably function as protein methyltransferases. Beside proteins there is a variety of different methyltransferase substrates including RNA, DNA, lipids and small molecules. Thus, methyltransferases are involved in diverse biological pathways (Martin and McMillan, 2002; Schubert et al., 2003). There are structurally defined types of S-adenosylmethionine dependent methyltransferases. Class I enzymes have a common seven-ß-strand structure and are most abundant. Protein arginine methyltransferases (PRMTs) and the lysine methyltransferase DOT1L fall into Class I. Protein lysine methyltransferases (PKMTs) belong in general to the SET [Su(var), Enhancer of zest, trithorax] domain superfamily (Class II) and catalyze methylation of protein lysine residues (Dillon et al., 2005; Petrossian and Clarke, 2011). In Figure 1 protein methyltransferases (PMT) are illustrated in a dendrogram based on sequence similarity. The nine PRMTs have similar sequences and cluster together. The much larger set of 56 PKMTs are more diverse. For a number PKMTs no activity has been verified so far.

1.1.1 Protein arginine methylation

The PRMTs transfer the methyl group from the AdoMet donor molecule to the acceptor molecule the terminal nitrogen atom of the guanidinium side chain of an individual arginine residue in the target protein (Bedford and Clarke, 2009; Gary and Clarke, 1998). Three distinct types of methylated arginine residues occur in mammalian cells: monomethylated arginine (MMA), symmetrically dimethylated arginine (sDMA) and asymmetrically dimethylated arginine (aDMA) (Figure 2) which in contrast to phosphorylation and acetylation create a supplementary level of information. These three derivates are present on a multitude of distinct protein species in the cytoplasm, nucleus and organelles of mammalian cells (Bedford and Clarke, 2009). According to the type of methylation catalyzed by the enzyme, the PRMTs were classified into different groups. While the type I PRMTs catalyze the formation of MMA and aDMA, the type II PRMTs form MMA and sDMA (Figure 2). PRMT1, PRMT3, CARM1 (PRMT4), PRMT6 and PRMT8 belong to the type I enzymes and PRMT5 and PRMT7 to the type II enzymes (Gary and Clarke, 1998; Krause et al., 2007; Pahlich et al., 2006).

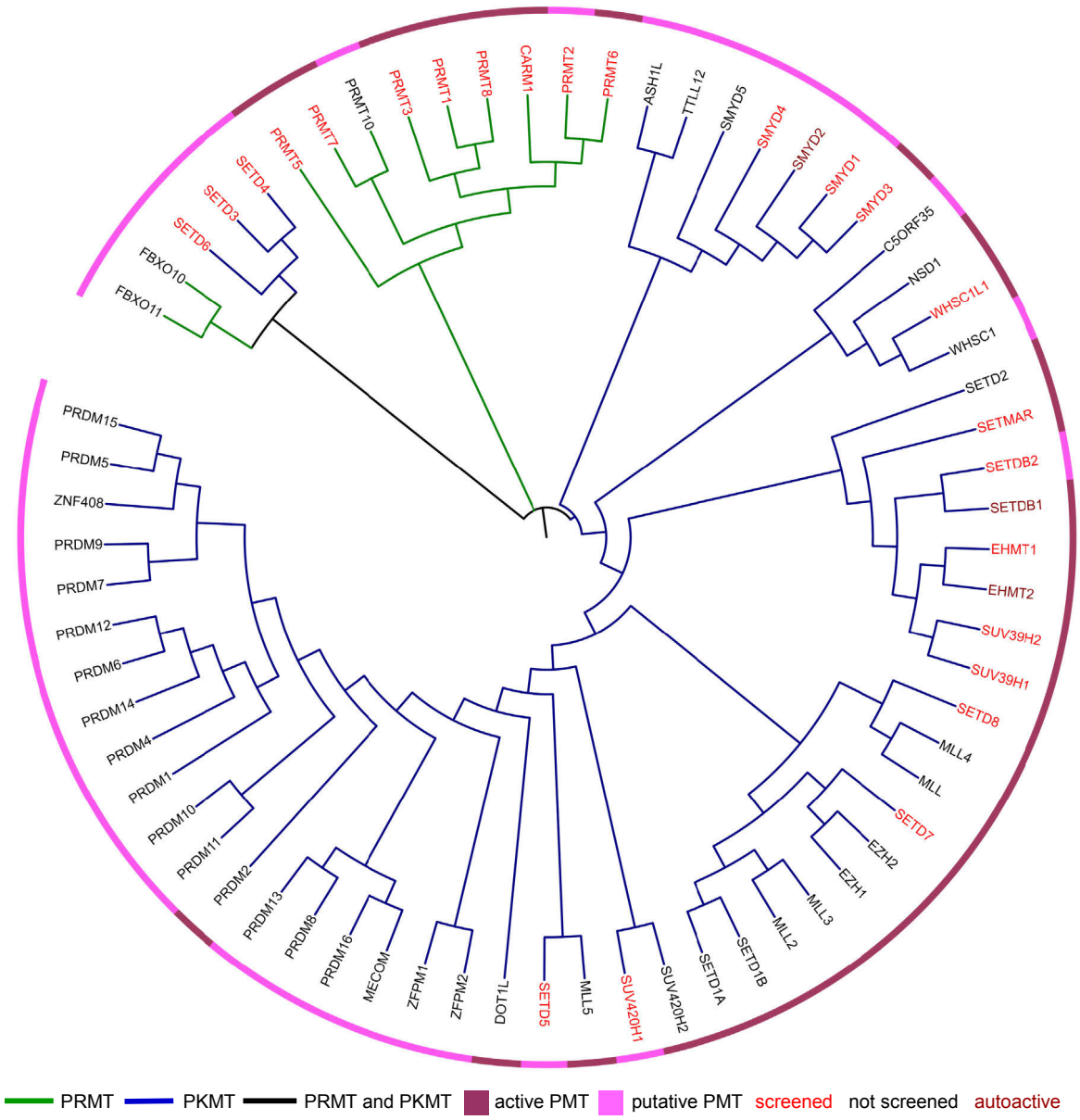

Figure 1: Dendrogram of protein methyltransferases

Protein methyltransferases (PMTs) are ordered based on a basic primary protein sequence alignment in CLUSTALW. The figure is generated using ITOL. Branches represent PRMTs (green), PKMTs (blue) and PRMT/PKMT (black). PMTs known to be active are indicated in dark purple and PMTs with unknown enzymatic activity are indicated in light purple. We will use a comprehensive set of PMTs to screen for protein-protein interactions. Label color indicates PMTs tested in our screens (red), not tested (black) or autoactive (dark red).

The seven PRMTs described above have been experimentally demonstrated to possess enzymatic activity while for PRMT2 and PRMT10 no activity has been demonstrated. PRMTs have a common catalytic methyltransferase domain which consists of a highly conserved core region and a subdomain important for binding of the methyldonor (Bachand, 2007). The individual PRMT family members differ in unique N-terminal regions of variable length and distinct domain motifs (Figure 3). PRMT2, PRMT3, PRMT8 and PRMT10 contain a Src homology 3 (SH3) domain, zinc finger (ZnF) domain, myristoylation (Myr) motif and tetratricopeptide repeats (TPR), respectively.

PRMT7 and PRMT10 have a second C-terminal catalytic domain.

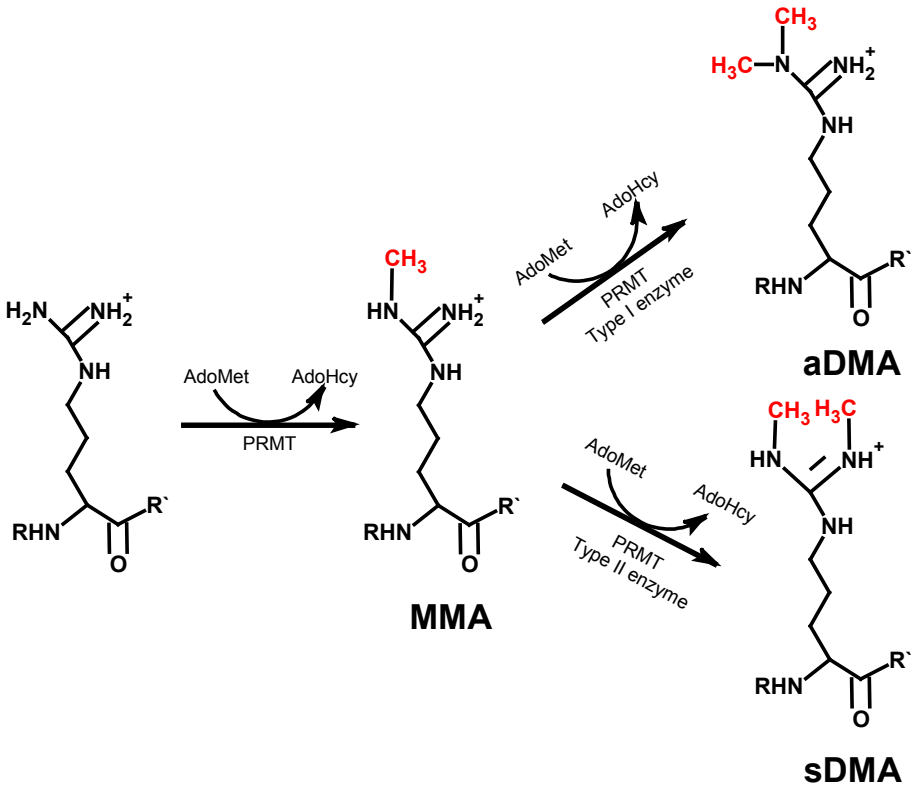

Figure 2: Types of methylation on arginine residues

Type I and II PRMTs generate monomethylarginine (MMA) on one of the terminal guanidino nitrogen atoms. The subsequent generation of asymmetric dimethylarginine (aDMA) is catalyzed by type I enzymes and the production of symmetric dimethylarginine (sDMA) is catalyze by type II enzymes. PRMTs use S-adenosylmethionine (AdoMet) as a methyldonor releasing S-adenosylhomocysteine (AdoHcy).

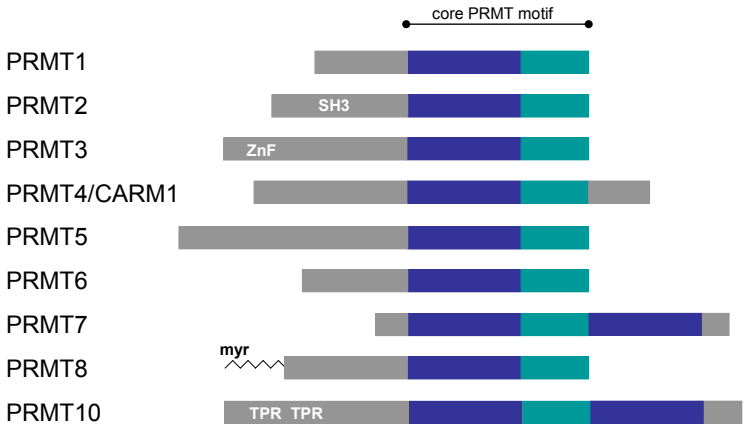

Figure 3: Protein arginine methyltransferases

The mammalian PRMT family currently contains nine highly related members. PRMTs have at least one core PRMT motif (blue). PRMT7 and PRMT10 have a duplication of this motif. PRMT2,

PRMT3, PRMT8 and PRMT10 contain a Src homology 3 (SH3) domain, zinc finger (ZnF) domain, myristoylation (Myr) motif and tetratricopeptide repeats (TPR), respectively.

1.1.1.1 PRMT1, the predominant type I methyltransferase, is closely related to PRMT8

PRMT1 was the first protein arginine methyltransferase in mammalian cells cloned and discovered independently by different groups (Abramovich et al., 1997; Lin et al., 1996). PRMT1 emerged as the predominant type I PRMT in mammalian cells contributing to over 85 % of type I protein arginine methyltransferase activity in cultured RAT1 fibroblast cells and mouse liver tissue (Pawlak et al., 2002; Tang et al., 2000a). PRMT1 knockout mice die shortly after implantation but embryonic stem cells generated from these mice are viable (Pawlak et al., 2000).

The first targets of PRMT1 identified were histones whose methylation is part of the histone code regulating gene expression (Abramovich et al., 1997; Lin et al., 1996). Later also non-histone proteins methylated by PRMT1 were identified. For example proteins involved in DNA damage response pathways as MRE11 and p53 binding protein 1 (53BP1) are methylated by PRMT1 in arginine-glycine (RG)-rich regions (Adams et al., 2005; Boisvert et al., 2005b). PRMT1 preferentially methylates RG-rich regions, a common feature of heterogeneous nuclear ribonucleoproteins (hnRNP) and other RNA binding proteins (RBP) that are the best known non-histone substrates of PRMT1 and are involved in various aspects of RNA processing and transport (Bedford and Richard, 2005; Liu and Dreyfuss, 1995). PRMT1 alters the subcellular localization of SYNCRIP, FUS and EWSR1 by methylation (Belyanskaya et al., 2001; Passos et al., 2006b; Rappsilber et al., 2003; Tradewell et al., 2012) but also affects protein-protein interactions (PPIs) (Bedford and Clarke, 2009; Bedford and Richard, 2005; McBride and Silver, 2001). For example, Sam68 contains proline-rich regions that interact with the SH3 and WW domains of several proteins. Interestingly, methylation decreases the binding of the SH3 domain but not WW domain allowing modulation of specific PPI (Bedford et al., 2000). Sam68 belongs to the family of signal transduction and activation of RNA (STAR) proteins. The STAR family proteins including SAM68, SLM-1, SLM-2, GRP33 and QKI-5 are known to be methylated. SAM68, SLM-2 and GRP33 are methylated by PRMT1 whilst the modifying enzyme of SLM-1 and QKI-5 is unknown (Cote et al., 2003).

PRMT8 was identified through its high degree of sequence identity to PRMT1. Although PRMT8 is closely related to PRMT1 it is expressed in specific tissues, especially in the brain, and is attached to the plasma membrane via N-terminal myristoylation (Lee et al., 2005d). PRMT8 activity is much lower when compared to PRMT1. However, removal of the elongated N-terminal of PRMT8 results in an enhanced enzymatic activity of PRMT8 suggesting that the N-terminal domain regulates PRMT8 activity. The N-terminal region of PRMT8 was detected to be automethylated on arginine 58 and 73 (Sayegh et al., 2007). In an pull-down experiment 20

PRMT8 binding proteins were identified including hnRNPs, RNA-helicases (DEAD box proteins), FUS , EWSR1, TAF(II)68 and caprin (Pahlich et al., 2008). Some of these proteins are also known to bind to PRMT1, including EWSR1 which is methylated by PRMT1 and PRMT8 (Kim et al., 2008). Structural analysis of PRMT1 revealed that it forms a dimer (Zhang and Cheng, 2003). PRMT8 was identified to interact with PRMT1. Hence, PRMT1 and PRMT8 form homo- and heterodimers (Kim et al., 2008; Lee et al., 2005b; Pahlich et al., 2008).

In summary, PRMT1 is involved in various processes including, signaling, DNA repair, transcriptional regulation, protein-protein interactions and localization. Most of the substrates identified are hnRNP proteins containing RG-rich regions but PRMT1 has been shown to methylated also non RG-rich proteins, suggesting that there are many more substrates which need to be identified.

1.1.1.2 CARM1 and PRMT6 are type I methyltransferases with pronounced substrate specificity

PRMT4 was discovered binding to the p160 family of nuclear receptor coactivators in a yeast-two hybrid (Y2H) approach. The association between PRMT4 and the p160 family enhances transcriptional activation as such PRMT4 is also called coactivator-associated arginine methyltransferase 1 (CARM1) (Chen et al., 1999). Embryos with a targeted distribution of CARM1 are small in size and die perinatally but embryonic stem cells generated from these CARM1-null embryos are viable (Kim et al., 2010; Yadav et al., 2003). In cells CARM1 forms a complex with ATP-remodeling (SWI/SNF) factors (Chi et al., 2004). CARM1 was identified to contribute to transcriptional regulation by methylation of histone 3 arginine 17 (Schurter et al., 2001). The first non-histone substrate identified was PABP1 a RNA binding protein (Lee and Bedford, 2002). Later on, substrates including transcriptional coactivators and a subset of splicing factors were identified (An et al., 2004; Cheng et al., 2007; Chevillard-Briet et al., 2002; Xu et al., 2001). The fact that CARM1 is a coactivator of nuclear receptors makes it a likely candidate for cancer (Bedford, 2007). In breast cancer tumors increased expression of CARM1 was shown (El Messaoudi et al., 2006). CARM1 was also involved in the development of prostate carcinomas (Hong et al., 2004). Recently, it has been shown that the carboxy-terminal domain (CTD) of RNAPII is methylated by CARM1 and that phosphorylation of the CTD inhibits methylation (Sims et al., 2011). An automethylation site on CARM1 regulates cellular functions but does not affect enzymatic activity of CARM1. Phosphorylation instead abolishes AdoMet binding and inhibits the methyltransferase activity of CARM1 (Feng et al., 2009; Kuhn et al., 2011). In summary, CARM1 catalyzes methylation of distinct substrates as PRMT1 and is involved in regulating several cellular processes including transcription, splicing and protein-protein interactions.

PRMT6 was discovered by searching the human genome for protein arginine methyltransferases. It is a nuclear enzyme with a substrate specificity functionally distinct from PRMT1 and CARM1. Like CARM1, PRMT6 displays an automethylation activity (Frankel et al.,

2002). Specific substrates of PRMT6 include the nuclear scaffold protein HMGA1a (Miranda et al., 2005), DNA Polymerase ß (El-Andaloussi et al., 2006), histone H3/H4 (Lee et al., 2004) and the HIV Tat protein (Boulanger et al., 2005; Invernizzi et al., 2007; Invernizzi et al., 2006; Xie et al., 2007). Hence, only two human non-histone proteins are known to be methylated by PRMT6.

1.1.1.3 PRMT5 is the major type II methyltransferase

PRMT5 was identified in a Y2H approach interacting with Janus kinase 2 (Jak2) and appears to be the major type II mammalian arginine methyltransferase (Branscombe et al., 2001). PRMT5 is conserved in eukaryotes and widely expressed in human tissues (Pollack et al., 1999). Like CARM1, PRMT5 can complex with hSWI/SNF ATP dependent chromatin remodeling protein and in this context it functions as a transcriptional coactivator (Dacwag et al., 2007). PRMT5 forms also with plCln a complex called the methylosome. This methylates the RG-rich regions of SmD1, SmD3 and SmB which is a prerequisite for the survival of motor neuron (SMN) dependent assembly of the spliceosomal core-UsnRNP (Brahms et al., 2001; Brahms et al., 2000; Friesen et al., 2002; Meister et al., 2001; Meister and Fischer, 2002). Additionally, PRMT5 methylates SPT5, p53 and FEN1 (Guo et al., 2010; Kwak et al., 2003). Recently crosstalk between arginine methylation and tyrosine phosphorylation was described on the epidermal growth factor receptor (EGFR) (Hsu et al., 2011). Hence, PRMT5 methylation crosstalks with other posttranslational modifications, regulates transcription, degradation, splicing, DNA replication and signaling in the cell.

1.1.1.4 Additional members of the PRMT family are not characterized

PRMT2 was identified by sequence homology to PRMT1 (Katsanis et al., 1997). PRMT2 transcripts are detected in most human tissues and are predominantly localized in the nucleus. PRMT2 null mice are viable and normal (Yoshimoto et al., 2006). A novel feature of PRMT2 is that it harbors a SH3 domain at its N-terminus (Scott et al., 1998) (Figure 3). Analysis using Y2H screening approaches identified PRMT2 interacting with the estrogens receptor alpha (Qi et al., 2002) and the androgen receptor (Meyer et al., 2007), both are nuclear hormone receptors. Although PRMT2 does not have enzymatic activity it does function as a coactivator for the estrogen receptor (Qi et al., 2002) and therefore is of special interest.

PRMT3 belongs to the type I enzymes and is expressed widely in human tissues with subcellular localization in the cytoplasm (Tang et al., 1998). Mouse embryos with a target disruption of PRMT3 are small size but survive after birth and attain a normal size in adulthood (Swiercz et al., 2007). An important feature of PRMT3 is the C2H2-type zinc-finger domain (Frankel and Clarke, 2000) (Figure 3). The ZnF domain of PRMT3 appeared to be necessary and sufficient for binding of rpS2. RpS2 was shown to be a bona fide *in vivo* substrate of PRMT3 that can not be modified by other PRMTs (Swiercz et al., 2005). Recently rpS2 was reported to be a substrate of Peptidylarginine deiminase 4 (PAD4, PADI4). PAD4 removes methylation by

conversion of arginine to citrulline, a process known as deimination (see 1.1.2) (Guo et al., 2011). PRMT3 has unique substrate specificity but only a handful of substrates are known.

PRMT7 was discovered by Miranda et al. and shown to monomethylate synthetic peptides (Miranda et al., 2004). Later it was shown that PRMT7 symmetrically dimethylates histones, myelin basic protein, fibrillarin and the spliceosomal protein SmB (Lee et al., 2005c). PRMT7 seems to be derived from a gene duplication event resulting in two putative substrate binding motifs (Figure 3). Both of these domains are required for the functionality of the enzyme as each separate domain was unable to function alone (Lee et al., 2005d). It is largely unknown which function PRMT7 has and no specific substrate is known.

PRMT10 was identified based on the homology to other PRMT family members as a product of a gene on the human chromosome 4y31 that is most closely related to PRMT7 (Cook et al., 2006; Krause et al., 2007; Lee and Stallcup, 2009). This gene is a candidate for PRMT10 and needs to be biochemically characterized. It contains two TPR which are known to mediate PPIs (Bedford et al., 2009).

The two F box-only family members, FBXO10 and FBXO11, have been suggested to be protein arginine methyltransferases but activity has not been clarified (Bedford and Clarke, 2009; Cook et al., 2006; Krause et al., 2007; Petrossian and Clarke, 2011).

1.1.2 Protein arginine demethylation

The JmjC enzyme JMJD6 was the first arginine demethylase identified. Using specific antibodies Chang et al. demonstrated that dimethylation of histone H3 and histone H4 peptides is reduced after incubation with JMJD6 (Chang et al., 2007). However, recent data showed that JMJD6 does not demethylate histone H4 and H3 fragment peptides but hydroxylates U2AF65 by incorporation of oxygen atoms (Webby et al., 2009). Lysyl hydroxylation is the dominant oxidative catalytic activity of JMJD6. JMJD6 catalyzed arginine demethylation cannot be ruled out but would have to occur at a level below current levels of detection (Webby et al., 2009).

The Peptidylarginine deiminase 4 (PAD4, PADI4) removes methylation by conversion of arginines to citrulline, a process known as deimination. PAD4 deiminates not only methylated arginine residues but also unmethylated arginine residues. Citrullination and methylation modifications are antagonistic to each other suggesting a conserved posttranslational regulatory strategy (Cuthbert et al., 2004; Guo et al., 2011; Wang et al., 2004). Peptidylarginine deiminases are not true “demethylases” as they do not convert monomethylarginine back to arginine. Whether arginine methylation can be removed or is a more static posttranslational modification that is removed via protein degradation, is a pivotal question in the field awaiting clarification.

1.1.3 Protein lysine methylation

Other than arginine, lysine residues are methylated on proteins. Furthermore, acetylation, ubiquitination, sumolation and neddylation are common modification on lysine residues (Yang and

Seto, 2008). The fact that alternated modifications occur on lysines provides another level of regulation since the presence of one inhibits the attachment of the other modification. Methylation can exist in mono-, di- and trimethylated state (Figure 4).

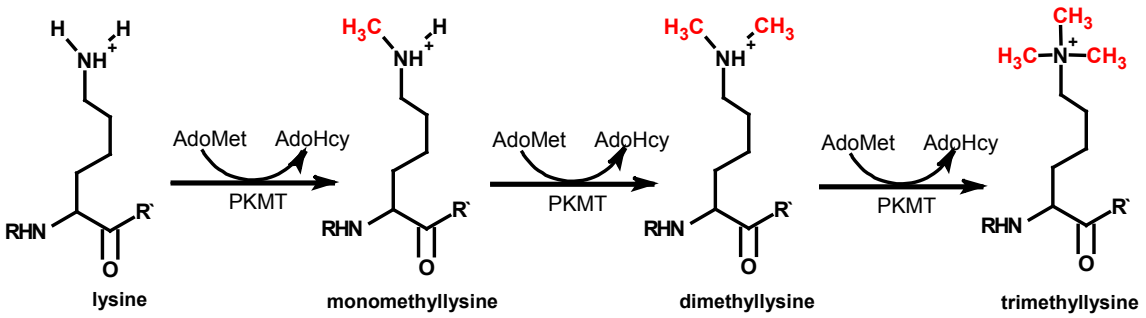

Figure 4: Types of methylation on lysine residues

PKMTs generate monomethyllysine, dimethyllysine and trimethllysine. PKMTs use S-adenosylmethionine (AdoMet) as a methyldonor releasing S-adenosylhomocysteine (AdoHcy).

Lysine methylation is accomplished by so called protein lysine methyltransferases (PKMT). All PKMTs, except Dot1, share a SET domain that is responsible for catalysis and binding of the cofactor AdoMet (Dillon et al., 2005). There are 56 PKMTs present in the human proteome but enzymatic activity and substrate specificity of most of them is unknown (Figure 1) (Levy et al., 2011; Petrossian and Clarke, 2011). At first lysine methylated histones were identified and later also non-histone proteins (Huang and Berger, 2008).

1.1.3.1 Chromatin state is regulated by histone lysine methylation

Histones play a dynamic role in controlling chromatin structure and transcription. There are different states of chromatin. The more compact chromatin is called heterochromatin and euchromatin, which is believed to be accessible for transcription, has a more open structure. Histone lysine methylation plays a major role in regulating the state of chromatin compaction. Additionally, histone lysine methylation regulates activation and repression of gene transcription within the euchromatin (Hublitz et al., 2009). In general histone H3 methylation at lysine 4, 36 and 79 correlates with euchromatin and transcriptional activation whereas histone H3 methylation at lysine 9 and 27 and histone H4 at lysine 20 is associated with heterochromatin and transcriptional repression (Bannister et al., 2002). However, different extend of methylation and crosstalks between different modifications leads to different functions in the cell (Yang and Seto, 2008) and thus reveals a “histone code” that extends the information of the genetic code (Jenuwein and Allis, 2001).

The first PKMT with enzymatic activity identified was SUV39H1. SUV39H1 specifically target histone H3 lysine 9 (H3K9me) (Rea et al., 2000). EHMT2 (G9a) mono- and dimethylates euchromatin at H3K9 (Tachibana et al., 2005). SUV39H1 trimethylates H3K9 and that correlates with the formation of heterochromatin (Robin et al., 2007). The trimethylated H3K9 is recognized

by heterochromatin protein 1 (HP1) through its chromodomain, but not the mono- and dimethylated H3K9. HP1 is localized to heterochromatin sites where it mediates gene silencing (Bannister et al., 2001; Jacobs and Khorasanizadeh, 2002). This shows that different levels of methylation can exert different functional outcomes. Although, histones are by far the predominant substrates described for PKMTs a few non-histone substrates are also known (Huang and Berger, 2008).

1.1.3.2 Non-histone protein functions are modulated by lysine methylation

There are a few non-histone substrates identified including cytochrome c in plants and fungi (Kluck et al., 2000; Polevoda et al., 2000). Calmodulin (Sitaramayya et al., 1980), Rubisco in plants (Houtz et al., 1989) and several ribosomal proteins in yeast were also found to be lysine methylated (Kruiswijk et al., 1978; Porras-Yakushi et al., 2005). The first human non-histone substrate identified was the tumor suppressor p53 which is methylated by SETD7 (SET9, SET7, SET7/9, KMT7). Methylation of p53 positively affects its stability and regulates the expression of p53 target genes (Chuikov et al., 2004). P53 is also methylated by SMYD2 on lysine 370 close to the lysine 372 modified by SETD7 (Huang et al., 2006). Monomethylation by SMYD2 has repressory function whereas dimethylation of lysine 370 leads to activation by association of p53 with the coactivator 53BP1. It was shown that the mono- and dimethylation mediated by SMYD2 can be removed by AOF2 (Huang et al., 2007). AOF2 also removes SET8 mediated lysine 382 methylation of p53 which leads to inactivation of p53 (Shi et al., 2007). Additionally, SETD7 methylates DNA methyltransferase 1 (DNMT1) (Esteve et al., 2009), TBP-associated factor TAF10 (Kouskouti et al., 2004), estrogen receptor (ER) α (Subramanian et al., 2008) and MYPT1 (Cho et al., 2011). A study by Dhayalan et al. applied a peptide array to determine an optimized target sequence for SETD7. Based on this they identified 91 new peptide substrates derived from human proteins and verified methylation of nine non-histone proteins (Dhayalan et al., 2011). The NFκB subunits RelA and p65 are methylated by SETD7 (Ea and Baltimore, 2009; Yang et al., 2009). SETD6 methylates RelA, too. SETD6 mediated methylation is recognized by EHMT1 which mediates H3K9 methylation and repression of NFκB signaling. Serine phosphorylation blocks the binding of EHMT1 and relieves repression of the target genes (Levy et al., 2011). The Vascular Endothelial Growth Factors Receptor 1 (VEGFR1) has been identified as the only methylation target of SMYD3 (Kunizaki et al., 2007). EHMT2 methylates non-histone proteins including CDYL1, WIZ, ACINUS and EHMT2 (automethylation). HP1 binds methylation specific to the automethylation site of EHMT2 and to the trimethylated peptides of histone H3, CDYL1 and WIZ. In addition, it was shown that methylation of CDYL1 abolished the interaction of the CDYL1 chromodomain to H3K9me3 (Rathert et al., 2008; Sampath et al., 2007).

In summary, most non-histone substrates have been identified methylated by SETD7 but also other PKMTs methylate non-histone proteins. The discussed cases indicate that lysine methylation of non-histone proteins can recruit methyl recognition domains, regulate stability of proteins,

repress function of proteins and plays a role in assembling of complexes. This implies a broader function of lysine methylation as previously thought. For a number of PKMTs no substrates have been verified so far and it is possible that these may methylate non-histone proteins.

1.1.4 Protein lysine demethylation

Two kinds of lysine demethylases have been identified: The lysine (K)-specific demethylase 1A (KDM1A, AOF2, LSD1) and the Jumonji C (JmjC) domain family proteins. AOF2 can demethylate mono- and dimethyllysine, but not trimethyllysine. AOF2 functions as histone demethylase (Shi et al., 2004) but is also involved in demethylating non-histone proteins including p53, MYPT1 and DNMT1 (Cho et al., 2011; Esteve et al., 2011; Shi et al., 2007; Wang et al., 2009). It is suggests that there are further non-histone targets demethylated by AOF2.

In 2006, the JmjC family protein JHDM1 was purified and shown to catalyze demethylation of lysine residues (Tsukada et al., 2006). Some of the JmjC family proteins have been shown to act on histones including JHDM2A, JMJD2A, JMJD2B, JMJD2C and JMJD2D (Whetstine et al., 2006; Yamane et al., 2006). Whilst no non-histone substrate has currently been described for the JmjC protein family, lysine methylation can at least in part, be described as a dynamic posttranslational modification that is regulated by demethylase and methyltransferase activities (Huang et al., 2007; Ruthenburg et al., 2007).

1.2 PMT substrate identification methods

1.2.1 Methods to detect methylated proteins

Arginine protein methylation has been implicated in various functions like protein-protein and protein-RNA interactions, cellular localization, nuclear transport, RNA processing, ribosome assembly, maturation of hnRNPs, translation accuracy, protein metabolism and cell signaling. (Choudhary et al., 2009; Kim et al., 2006a). There is a large family of protein methyltransferases which are expressed in a large variety of tissues and diverse subcellular localization, predominantly nuclear but also strictly cytoplasmatic, like PRMT5 (Herrmann et al., 2009). However, methyl binding domains containing proteins exist (Taverna et al., 2007) suggesting that protein methylation plays a ubiquitous role in many cellular processes other than epigenetic gene regulation. Progress to date has undoubtedly uncovered only a small portion of the roles of arginine methylation in regulation of protein function in biological processes (Lee and Stallcup, 2009). Likewise, lysine methylation of proteins regulate diverse cellular processes and different publications suggest that there will be far more non-histone proteins discovered to be lysine methylated (Huang and Berger, 2008). Similar to protein acetylation that was initially characterized primarily on histones but later recognized as modification on many non-histone proteins (Gu and Roeder, 1997; Lee et al., 2005a; Sterner and Berger, 2000). Proteome-wide studies revealed that over 1700 proteins are acetylated. To understand the biological role of methylation, besides histone

methylation, proteome-wide screens are required to identify new lysine methylation substrates and the responsible enzymes. It is conceivable that there are more non-histone substrates and additional PKMTs methylating non-histone targets.

The detection of non-histone targets has proved technically challenging. Different immunopurification (IP) approaches coupled to mass spectrometry have been implicated to identify methylated proteins (Boisvert et al., 2003; Ong et al., 2004). Those studies are essentially limited by the quality of available antibodies. Many commercial available methyl antibodies crossreact with unmodified sequences or they fail to efficiently bind a large variety of methylated sequences (Levy et al., 2011). Further limitation of the IP approach is that the corresponding enzyme is not identified (Komyod et al., 2005; Pahlich et al., 2006).

Another approach to identify non-histone targets is to query the proteome for a linear amino acids sequence which surrounds the methylation sites. To do so, a specific profile of the protein methyltransferase methylation site is determined by methylation of peptide arrays. This specific profile is used in a proteome-wide searched for similar motifs. A proteome-wide search for EHMT2 specific profiles revealed 92 human proteins and showed methylation on eight in a direct *in vitro* methylation assay (Rathert et al., 2008). However, the motif search works only for enzymes recognizing linear amino acid sequences which might not be the case for all PKMTs and PRMTs.

There are two unbiased screening approaches to identify methylation substrates. First, hypomethylated cell extract of PC12 rat cells can be used to visualize methyl acceptor proteins. In this approach the hypermethylated cell extract is incubated with the radiolabel AdoMet in the presence of protein translation inhibitor. Proteins incorporate the radioactive labeled methyl group and were separated by 2D gel electrophoresis. More than 50 different methyl acceptor proteins were detected by fluorography. However, identification of these proteins was not possible (Liu and Dreyfuss, 1995; Najbauer and Aswad, 1990; Pahlich et al., 2006; Sampath et al., 2007). In a similar approach numerous heterogeneous nuclear ribonucleoproteins (hnRNPs) were discovered by immunopurification and 2D gel electrophoresis (Najbauer and Aswad, 1990; Pahlich et al., 2006). Those approaches are limited by the detection and identification of low abundant proteins.

An alternative approach is the human protein microarray based platform which is incubated with the appropriate PKMT and AdoMet. Levy et al. identified 216 potential SETD6 substrates and showed methylation on six in a direct *in vitro* methylation assay (Levy et al., 2011). The readout was performed by methyl specific antibodies or radioactive labeled AdoMet. Antibodies are difficult to use for readout as discussed above. Radioactive labeling has inherent limitations in its signal to noise ratio and imaging of radioactivity exposed on film. In addition, the protein array system has several limitations. First, only one third of the human proteome is represented on the protein array. Second, many proteins on the protein array do not cover the full length sequence of the protein. Third, signal intensity of the array is low (Levy et al., 2011).

In summary, there are proteome-wide approaches to detect protein methylation but all have strong limitations in discovering methylated proteins.

1.2.2 The Y2H system to find PMT interacting proteins and potential substrates

A physical interaction is a requisite for PMT enzymes to methylate target proteins. Therefore, in principle substrates of PMTs ought to be identified by studying PMT-protein interactions (Passos et al., 2006a). The analysis of protein interactions is based on two main technologies: The Y2H system to detect binary interactions and affinity purification followed by mass spectrometry (AP-MS) which detects direct and indirect interactions. In an AP-MS approach stable protein complexes composed around the protein of interest are isolated from cells and then analyzed using mass spectrometry (Gingras et al., 2007). Both approaches have advantages and disadvantages. The analysis of human PPIs in the Y2H system provides a heterologous environment whereas the AP-MS approach enables to study complexes that assemble in a more natural cellular environment (Choudhary and Mann, 2010). Washing steps during affinity purification leads to the loss of weak or transiently bound proteins and hence these interactions are not detected in the mass spectrometer. Highly abundant proteins are predominantly identified in the AP-MS approach, often also as false positive interaction partners. The Y2H approach assays proteins independent of endogenous expression levels. Proteins are expressed at very low levels and tested systematically, thus can be statistically analyzed (Worseck et al., 2012). Recent studies proved that large scale Y2H analyses result in high-precision data. A study by Yu et al. demonstrated that Y2H and AP-MS data are of equal high-quality if taken into account the different and complementary nature of the interactions determined. They also showed that the binary interaction map in comparison to the co-complex interactome model is enriched for transient signaling interactions (Simonis et al., 2009; Venkatesan et al., 2009; Yu et al., 2008). It was also shown that direct human PPIs relevant for signal transduction are identified by systematic Y2H interaction screening (Vinayagam et al., 2011). Enzyme-substrate interactions are direct and often transient. Therefore, the Y2H system is an efficient method to detect PMT-protein interactions (Passos et al., 2006a) and additionally it is independent of affinity tools that are lacking for the analysis of protein methylation.

Y2H system to detect PPIs. The Y2H system was originally developed by Stanley Fields (Fields and Song, 1989). It is based on two separable domains of a transcription factor, the DNA-binding domain (DBD) that recruits the transcription factor to the DNA and the activation domain (AD) that initiates transcription of target genes. In the Y2H system, those two domains are separately fused to the proteins of interest resulting in DBD and AD fusion proteins referred to as bait and prey protein, respectively. An interaction between the fusion proteins in yeast cells expressing both hybrid proteins brings the DBD and the AD into close proximity and leads to reconstitution of the transcription factor (Figure 5). The functional transcription factor activates

the expression of one or more reporter genes. This enables the growth selection of yeast cells harboring an interacting protein pair (Figure 5).

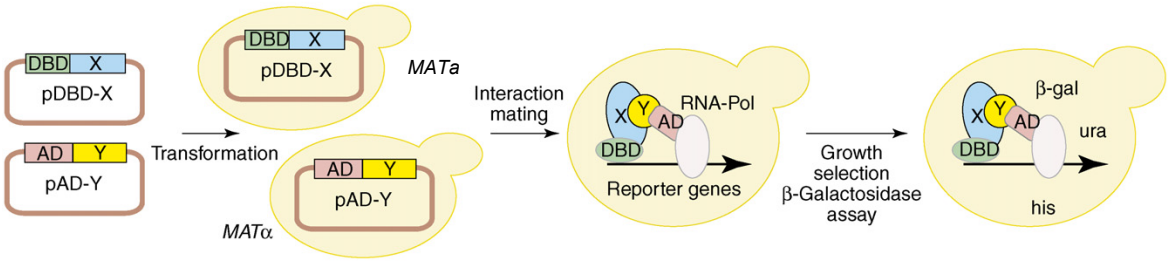

Figure 5: Principle of the Y2H system for the detection of binary interactions

Coding sequences for a protein X and a protein Y are fused to a DNA-binding domain (DBD, i.e. bait plasmid) and a transcription activation domain (AD, i.e. prey plasmid). Two yeast strains of opposite mating type MATa and MATα are transformed with the bait and prey plasmids, respectively and mated. The diploid yeast is expressing both hybrid proteins. Upon interaction of protein X and protein Y, transcriptional activity of the DBD and the AD-domains is reconstituted leading to reporter gene activation. Reporter gene expression enables yeast growth in the absence of selected nutrients, such as histidine and uracil. In many Y2H systems, the lacZ gene is also utilized as reporter, so that an interaction of protein X and protein Y can be assayed via b-galactosidase activity (Stelzl and Wanker, 2006).

There are two Y2H high-throughput (HTP) methods: The matrix and the library approach. In the library approach, baits are screened separately against a prey pool (the prey library) containing random cDNA fragments or open reading frames (ORFs) (Chien et al., 1991). Yeast cells containing a positive interaction pair of proteins are selected based on the reporter gene activation and the resulting ability to grow on selective medium. In the Y2H library screen the plasmid DNA of the interacting prey proteins are separately isolated form the yeast colonies growing on the selective medium and are identified by Sanger sequencing (Chien et al., 1991). On the contrary, in the Y2H-matrix screen the prey clones are well characterized. In each position of array a particular prey is expressed. The baits are screened separately against the prey array and interacting protein pairs are identified based on the expression of the reporter gene resulting in growth on the selective medium in a specific position on the array. Hence, the identity of the prey protein can be identified via its position in the array (Worseck et al., 2012). The proteins encoded by the prey clones used in the Y2H-matirx are tested with the equal probability in contrast to the random protein encoding prey clones in the Y2H library screen which are not normalized. Hence, interactions are detected much more efficiently in the Y2H-matrix screen using a set of normalized preys (Reboul et al., 2003). Additionally, Y2H-matrix screens can be repeated (Worseck et al., 2012).

The relative low cost and the possibility for automation make HTP Y2H experiments to one of the most powerful tools to generate quality-controlled, proteome-wide, binary PPI maps, such as those generate for yeast (Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008), fly (Giot et al., 2003), worm (Li et al., 2004; Simonis et al., 2009) and human (Bandyopadhyay et al., 2010; Rual et al., 2005; Stelzl et al., 2005; Vinayagam et al., 2011; Yu et al., 2011). Limitations of the Y2H-matrix

approach will be addressed in this study that describes a novel Y2H screening approach that utilizes second generation sequencing.

1.3 Aim of this study

Methylation is a post translational modification that can regulate diverse cellular processes and is subjected to increasing investigation. Several lines of evidence are suggesting that protein methylation plays a ubiquitous role in many cellular processes other than epigenetic regulation. There are around 100 proteins known to be methylated on arginine residues but only for around 50 the responsible PMT is known. Even fewer lysine methylated proteins are known and so far there is only a handful of PKMTs known to methylate non-histone proteins. For a large set of PKMTs no methylation substrates has been identified so far. Despite the importance of methylation in maintaining cellular homeostasis, the development of proteome-wide approaches for detecting this modification has been limited and proven technically challenging. The aim of this study was to develop a proteome-wide methode to identify proteins interacting with PMTs or demethylases, assuming that a large fraction of identified proteins are methylation substrates of PMTs.

The HTP Y2H-matrix approach produces binary, high-quality PPI data and is an efficient method to detect the often transient enzyme-substrate interactions. However, the sensitivity of the screening method and thus low data coverage is a severe limitation. Hence, we developed a workload reduced Y2H approach with significantly increased sensitivity. The novel method utilized a second generation sequencing step and thus abbreviated Y2H-seq. As transient PPIs are efficiently detected if the sampling is increased the high sampling sensitivity of the Y2H-seq approach is of particular importance when detecting transient enzyme-substrate interactions, such as those between methyltransferases and their substrates.

The novel Y2H-seq approach was evaluated against our state of art Y2H-matrix screening approach and used to screen 8 PRMTs, 17 PKMTs and 10 PDeMs against a set of 14,268 proteins. We generated a high-quality interaction network consisting of 523 interactions with 324 prey proteins. This is the first proteome-wide interaction dataset of enzymes involved in methylation. Despite the few proteins known to be methylated and the incompleteness of the prey array we identified 11 prey protein substrates already known to be methylated. Whilst not all interacting proteins will be methyltransferase substrates it is clear methylation substrates of PMTs can be identified in the Y2H approach. Hence, methylation substrates of PMTs can be identified in the Y2H approach. Therefore, the interaction network will serve as resource to identify new methylation substrates. We developed a methylation assay using radioactive methyl donor and identified SYNCRIP to be methylated. To identify methylation sites an *in vitro* methylation assay was established and coupled to mass spectrometry analysis. Thus, new candidate methylation substrates were validated and the sites of methylation identified. On seven of nine proteins tested methylation sites were mapped including SPIN2B, DNAJA3, QKI, SAMD3, OFCC1, SYNCRIP

and WDR42A.

2 Material

2.1 Chemicals

4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Sigma-Aldrich, Taufkirchen)
2-mercaptoethanol (Merck, Darmstadt)
Acetic acid (Merck, Darmstadt)
Acrylamid/Bisacrylamid 40 % (37,5:1) (Roth, Karlsruhe)
Adenosine-5'-triphosphate (ATP) (Sigma-Aldrich, Taufkirchen)
Adenosyl-L-methionine, S-[methyl-3H]; (AdoMet[3H]) (PerkinElmer, USA)
Agarose (Sigma-Aldrich, Taufkirchen)
Ammonium persulfate (APS) (Merck, Darmstadt)
Ammonium sulfate (Merck, Darmstadt)
Ampicillin trihydrate (Sigma, Deisenhofen)
Bacto Yeast Extract (BD BIOSCIENCES, USA)
Bacto-Agar (BD BIOSCIENCES, USA)
Bacto-Pepton (BD BIOSCIENCES, USA)
Bacto-Tryptone (BD BIOSCIENCES, USA)
Betain (Sigma-Aldrich, Taufkirchen)
Boric acid (Merck, Darmstadt)
Bovine serum albumin fraction V (BSA) (Roche, Mannheim)
Bromphenol blue (Merck, Darmstadt)
Calcium chloride dihydrate (Merck, Darmstadt)
Chloramphenicol (Boehringer, Mannheim)
Chloroform (Merck, Darmstadt)
Coomassie Brilliant Blue G-250 (Biomol GmbH, Hamburg)
Deoxyribonucleic acid sodium salt from salmon testes (Sigma-Aldrich, Taufkirchen)
Dimethyl sulfoxide 99.9 % (DMSO) (Sigma-Aldrich, Taufkirchen)
Dipotassium phosphate (Acros organics part of Thermo Fisher Scientific Inc., Geel, Belgium)
Dithiothreitol (DTT) (Roth, Karlsruhe)
D-luciferin sodium lyophilized firefly (Sigma-Aldrich, Taufkirchen)
Dulbecco’s modified eagle medium (DMEM+GlutaMAX™-I) (Gibco BRL, Gaithersburg, USA)
Dulbecco’s phosphate buffered saline (DPBS) (Gibco BRL, Gaithersburg, USA)
Ethanol (Merck, Darmstadt)
Ethylene glycol tetraacetic acid (EGTA) (Roth, Karlsruhe)
Ethylenediaminetetraacetic acid (EDTA) (Roth, Karlsruhe)
Fetal bovine serum (FBS) (qualified FBS, south american) (Gibco BRL, Gaithersburg, USA)
Glucose monohydrate (Merck, Darmstadt)

Glycerol (Merck, Darmstadt)

Glycine (MP Biochemicals, Aurora, USA)

Glycogen (Roche, Mannheim)

Glycylglycine (Acros organics part of Thermo Fisher Scientific Inc., Geel, Belgium)

Histidine (Sigma-Aldrich, Taufkirchen)

Imidazole (Sigma-Aldrich, Taufkirchen)

Isopropanol (Merck, Darmstadt)

Isopropyl-ß-D-thiogalactopyranosid (IPTG), diaxone-free (Fermentas GmbH, St. Leon-Rot)

Kanamycin sulfate (Sigma-Aldrich, Taufkirchen)

Leucin (Sigma-Aldrich, Taufkirchen)

Lithiumacetate (LiOAc) (Sigma-Aldrich, Taufkirchen)

Magnesium chloride (Roth, Karlsruhe)

Magnesium sulfate (Roth, Karlsruhe)

Methanol (Merck, Darmstadt)

Monopotassium phosphate (Roth, Karlsruhe)

Opti-MEM I (Gibco BRL, Gaithersburg, USA)

Orthophosphoric acid (Alfa Aesar GmbH & Co KG, Karlsruhe)

Paraformaldehyde (PFA) (Roth, Karlsruhe)

Phenol (Roth, Karlsruhe)

Phenylmethansulfonylfluorid (PMSF) (Roth, Karlsruhe)

Polyethylene glycol (PEG) 3350 (Sigma-Aldrich, Taufkirchen)

Polyethylene glycol (PEG) 8000 (Sigma-Aldrich, Taufkirchen)

Potassium acetate (Merck, Darmstadt)

Potassium chloride (Roth, Karlsruhe)

Protease inhibitor (Roche, Mannheim)

S-(5'-Adenosyl)-L-methionine chloride (AdoMet) (Sigma-Aldrich, Taufkirchen)

Sodium carbonate (Merck, Darmstadt)

Sodium chloride (Roth, Karlsruhe)

Sodium citrat (Roth, Karlsruhe)

Sodium dihydrogen phosphate (Merck, Darmstadt)

Sodium dodecyl sulfate (SDS) (Roth, Karlsruhe)

Sodium hydrogencarbonate (Merck, Darmstadt)

Sodium hydroxide (Roth, Karlsruhe)

Sorbitol (Sigma-Aldrich, Taufkirchen)

Spectinomycin dihydrochloride pentahydrate (Sigma-Aldrich, Taufkirchen)

Sucrose (Merck, Darmstadt)

SYBR Gold Nucleic Acid Gel Stain (Invitrogen, Darmstadt)

Tetracycline hydrochloride (Sigma-Aldrich, Taufkirchen)

Tetramethylethylenediamine (TEMED) (Invitrogen, Darmstadt)
Thiamine hydrochloride pure (AppliChem, Darmstadt)
Tris (hydroxymethyl) aminomethane (Tris Base) (Roth, Karlsruhe)
Tris (hydroxymethyl) aminomethane hydrochloride (Tris HCl) (Sigma-Aldrich, Taufkirchen)
Triton X-100 (Sigma-Aldrich, Taufkirchen)
Tryptophan (Sigma-Aldrich, Taufkirchen)
Tween 20 (Sigma-Aldrich, Taufkirchen)
Uracil (Sigma-Aldrich, Taufkirchen)
Yeast nitrogen base (Difco part of BD Biosciences, USA)

2.2 Lab ware

384-well MTPs, PS, flat bottom, clear, sterile, with lid (Greiner Bio-One GmbH, 781186)
96-well deepwell plates (2000 µl/well) (Eppendorf, 0030 501.322)
96-well MTPs, PS, flat bottom, crystal clear (Greiner Bio-One GmbH, 655101)
96-well MTPs, PS, flat bottom, lumnitrac600, high binding, white, sterile (Greiner Bio-One GmbH, 655074)
96-well MTPs, PS, flat bottom, TC, white, sterile (Greiner Bio-One GmbH, 655073)
96-well PCR plate (Costar part of Corning Incorporated, 6511)
Agar-plates (241 x 241 x 20) (Nunc GmbH & Co. KG, 240845)
BiomekNX (Beckman Coulter GmbH)
Biophotometer Plus (Eppendorf AG)
Centrifuge 5810 R (Eppendorf AG)
E.A.S.Y 429k digital camera (Herolab GmbH Laborgeräte)
Fixed-angle rotors F-45-30-11 (Eppendorf AG)
Glass beads, acid-washed 425-600 µm (Sigma-Aldrich, G8772)
Incubator 1000 (Heidolph Instruments GmbH & Co. KG)
InfiniteM200 multimode microplate reader (Tecan Group Ltd.)
Innova44 shaker (New Brunswick Scientific)
Kby roboter (Cambridge, UK)
MicroPulser electroporation apparatus (Bio-Rad Laboratories)
Mini-PROTEAN tetra cell electrophoresis system (Bio-Rad Laboratories)
NanoDrop ND-1000 (Thermo Fischer Scientific Inc.)
Ni-NTAagarose beads (Qiagen GmbH, Hilden)
Nitrocellulose Membrane (Bio-Rad Laboratories, 162-0115)
Omnitrays (Nunc GmbH & Co. KG, 165218)
Pin tools with 96 and 384 pins. The steel pins are cylindrical with a diameter of 1.3 mm and the edge of the flat top that is touching the agar is bevelled 45°at 0.2 mm. Sterilize by heating the pins until they glow red. Let them cool in a sterile environment.

Plastic tape for sealing PCR plates (Costar; 6524 or Thermo Fischer Scientific Inc., AB-5558)
Polyvinylidene fluoride (PVDF) membrane (Bio-Rad Laboratories, 162-0177)
PowerPac Universal Power Supply (Bio-Rad Laboratories)
Sonifier B-12 Cell Disruptor (Branson Sonic Power Company, Danbury, USA)
Sterile breathable sealing films (Aeraseal, Excel Scientific Inc., BS-25)
Sunrise 96 horizontal gel electrophoresis apparatus (Biometra GmbH)
Swing-bucket rotor A-4-81 (Eppendorf AG)
Tetrad PTC-225 thermo cycler (MJ Research Inc.)
Thermomixer comfort (Eppendorf AG)
Tissue culture flask (TPP Techno Plastic Products AG, 90076)
Tissue culture dish (TPP Techno Plastic Products AG, 93100)
Titramax 1000 (Heidolph Instruments GmbH & Co. KG)
Trans-Blot SD Semi-Dry transfer cell (Bio-Rad Laboratories)

2.3 Enzyme, proteins, DNA, kits

1 Kb Plus DNA ladder (Invitrogen, USA)
AttoPhos substrate set (Roche, Mannheim)
Benzonase (Sigma-Aldrich, Taufkirchen)
BP Clonase Enzyme Mix (Invitrogen, USA)
Bright-glo luciferase assay system (Promega, Madison)
Coenzyme A (CoA) (Sigma-Aldrich, Taufkirchen)
dNTP-Mix (Fermentas GmbH, St. Leon-Rot)
FastDigest Bsp1407I (Fermentas GmbH, St. Leon-Rot)
Lipofectamine 2000 (Invitrogen, USA)
LR Clonase Enzyme Mix II (Invitrogen, USA)
Lysozym (Roth, Karlsruhe)
PfuTurbo DNA Polymerase (Stratagene, Santa Clara)
Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes, Vantaa)
Prestained protein ladder, PageRuler™ Plus (Fermentas GmbH, St. Leon-Rot)
Proteinase K solution (Invitrogen, USA)
PureYield Plasmid Midiprep System (Promega, Madison)
QIAprep spin miniprep kit (Qiagen GmbH, Hilden)
QIAquick PCR purification Kit (Qiagen GmbH, Hilden)
QuikChange site-directed mutagenesis kit (Stratagene, Santa Clara)
Salmon sperm DNA (carrier DNA) (Sigma-Aldrich, Taufkirchen)
Trypsin-EDTA (Gibco BRL, Gaithersburg, USA)
Western lightning plus-ECL (PerkinElmer, Massachusetts)
Zymolase 20T (Seikagaku Corporation)

2.4 Organism

2.4.1 Bacteria strains

DH10B: F`*mcrA* Δ-(*mrr hsd RMS-mcr BC*) φ80d*lacZ*ΔM15 Δ*lacX*74 *deoR recA*1 *araD*139 Δ(*ara leu*)7697 *galU galK* λ⁻ *rpsL endA*1 *nupG* (Invitrogen)

XL1-Blue: *recA*1 *endA*1 *gyrA*96 *thi*-1 *hsdR*17 *supE*44 *relA*1 *lac*⁻ F`[*pro*AB *lac*lqZΔM15 Tn10 (Tet^r) (Stratagene)

SCS1: *hsd*R17(rK mK +) *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1 *sup*E44 carrying the pRARE plasmid (Stratagene)

2.4.2 Yeast strains

L40ccU: MAT**a** his 3Δ200 trp1-901 leu 2-3,112 LYS 2 :: (lex Aop)4--HIS3 ura3:: (lex Aop)8 -lacZ ADE 2 :: (lexAop)8-URA3 GAL4 gal 80 could 1 cyh 2 (Goehler et al., 2004)

L40cc: MATα his3Δ200 trp1-910 leu2-3,112 ade2 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ GAL4 gal80 can1 cyh2 (Goehler et al., 2004)

2.4.3 Mammalian cell lines

The HEK 293 cell line is a permanent line of primary human embryonic kidney transformed by sheared human adenovirus type-5 (Ad 5) DNA. HEK293 are adherent fibroblastoid cells growing as a monolayer.

2.5 Reagents

2.5.1 E. coli and yeast miniprep

Buffer P1
50 mM Tris pH 8.0
10 mM EDTA pH 8.0
add 50 mg/l RNAse A after autoclaving

Buffer P2
0.2 M NaOH
1 % SDS
Buffer P3 pH 5.5
3 M Potassium acetate

SCE-Buffer

1 M Sorbitol

0.1 M Sodium citrat pH 5.8

10 mM EDTA pH 8.0

Buffer H1

1 M Sorbitol

0.1 M Sodium citrat pH 5.8

10 mM EDTA pH 8.0

1 mg/ml Zymolase 20T

0.01 M DTT

add 50 mg/l RNAse A

2.5.2 Yeast transformation

10x Tris/EDTA buffer (10x TE) pH 7.5

0.1 M Tris pH 7.5

10 mM EDTA pH 8.0

Carrier DNA

1x TE

5 mg/ml Salmon sperm DNA (Sigma)

Mix1

0.5x TE

0.1 M LiOAc

1 M Sorbitol

Mix 2

1x TE

0.1 M LiOAc

40 % PEG-3350

2.5.3 Agarose gel electrophoresis

10× TBE pH 8,3

0.89 M Tris

0.89 M Boric acid

0.02 M EDTA pH 8.0

10× Orange G Sample buffer
50 % sucrose
0.5 % Orange-G

2.5.4 SDS polyacrylamide gel electrophoresis and western blot

4x SDS gel loading buffer
0.2 M Tris pH 6.8
4 % SDS
40 % Glycerine
0.4 % Bromphenol blue
prior to use add 0.2 mM DTT

10x Electrophoresis buffer
0.25 M Tris Base
1.92 M Glycine
1 % SDS

10x stock blotting buffer
0.25 M Tris-Base
0.2 M Glycine
0.375 % SDS

Transfer buffer
1x Stock blotting buffer
20 % Methanol

10xTBS pH 7.5
0.25 M Tris Base
0.15 M NaCl

TBST
1x TBS
0.01 % Tween 20

Blocking buffer
1x TBST
5 % BSA (Bovine Serum Albumine Fraction V)

Coomassie-Blue stain

50 % Methanol

10 % Acetic acid

0.175 % Coomassie Brilliant Blue G-250

Destain

30 % Methanol

10 % Acetic acid

Blue silver stain

20 % Methanol

10 % Phosphoric acid

10 % Ammonium sulfate

0.12 % Coomassie Brilliant Blue G-250

2.5.5 Co-immunoprecipitation

Hepes-buffer

50 mM Hepes pH 7.4

150 mM NaCl

1 mM EDTA

10 % Glycerin

1 % Triton X-100

Protease inhibitor (Roche 11051600)

Luciferase substrate

0.25 M Glycylglycine

0.15 M KxPO4 pH 8.0

40 mM EGTA

20 mM ATP

10 mM DTT

0.15 M $MgSO_4$

1 mM CoA

75 µM Luciferin

0.5M KxPO4 pH 8.0

9.4 ml 0.5M K_2HPO4

0.6 ml 0.5M KH_2PO4

1x TBST II

10 mM Tris-Base

0.15 M NaCl

0.05 % Tween-20

1x Carbonate buffer

70 mM $NaHCO_3$

30 mM Na_2CO_3

2.5.6 E. coli protein expression and purification

NiNTA Resuspension buffer

20 mM Tris-HCl pH 7.5

0.3 M NaCl

0,2mg/ml Lysozyme

20 mM Imidazole

Protease Inhibitor tablet

Lysozyme stock (stored at -20 °C)

20 mM Tris pH 7.5

150 mM NaCl

10 mg/ml Lysozyme

Wash buffer

20 mM Tris-HCl, pH 7.5

0.3 M NaCl

20 mM Imidazole

Ni-NTA agaraose beads

Prepare 50 % slurry equilibrated in wash-buffer

2.6 Media

2.6.1 E. coli growth media

LB-medium pH 7.2

10 g/l Bacto tryptone

5 g/l Yeast extract

10 g/l NaCl

LB agar pH 7.2

10 g/l Bacto tryptone

5 g/l Yeast extract

10 g/l NaCl

20 g/l Agar

SOB-Medium

20 g/l Tryptone

5 g/l Yeast extract

0.5 g/l NaCl

add after autoclaving:

10 mM $MgCl_2$

10 mM $MgSO_4$

SOC-Medium

SOB-Medium

4g/l Glucose

2YT-Medium pH 7,2

16 g/l Bacto tryptone

10 g/l Yeast extract

5 g/l NaCl

Transformation and Storage Solution (TSS)

85 % LB-medium

10 % (w/v) PEG 8000

5 % DMSO

50 mM MgCl2

Filter sterilize through a 0.22 µm pore filter

SB Medium

12 g/l Bacto tryptone

24 g/l Yeast extract

4 % Glycerol

Antibiotics concentration

100 µg/ml Ampicillin

20 µg/ml Tetracycline

15 µg/ml Kanamycin
50 µg/ml Spectinomycin

2.6.2 S. cerevisiae growth media

YPD liquid medium
10 g/l Yeast extract
20 g/l Peptone

YPD agar
10 g/l Yeast extract
20 g/l Peptone
20 g/l Agar

Yeast liquid medium (NB)
6.7 g/l Yeast nitrogen base

Agar
20 g/l Agar

Yeast storage medium (NBG)
6.7 g/l Yeast nitrogen base
25 % Glycerol
58.88 g/l Betain

Glucose stock solution
400 g/l Glucose monohydrate

Amino acid concentration
100 mg/l Leucin
20 mg/l Histidine
20 mg/l Uracil
20 mg/l Tryptophan
20 mg/l Adenine

2.6.3 Mammalian cell culture media

Cell culture medium
Dulbecco's modified eagle medium (DMEM+GlutaMAX™-I)
10 % FCS (fetal bovine serum)

2.7 Vectors

2.7.1 Gateway entry vectors

pDONR221

Size: 4762 bp

Negative selection: ccdB

Bacterial resistance: Kanamycin

Reference: Invitrogen

pDONR223

Size: 5005 bp

Negative selection: ccdB

Bacterial resistance: Spectomycin

Reference: Invitrogen

2.7.2 Expression vectors for E. coli

pDESTco

Size: 7062 bp

Fusion protein or tag: Polyhistidine

Negative selection: ccdB

Bacterial resistance: Ampicillin

Reference: Scheich et al, 2007

pDESTcoG

Size: 7698 bp

Fusion protein or tag: Glutathion S-ransferase

Negative selection: ccdB

Bacterial resistance: Ampicillin

Reference: Scheich et al, 2007

pRARE

Size: 4694

Bacterial resistance: Chloramphenicol

Reference: Novy et al, 2001

2.7.3 Expression vectors for S. cerevisiae

pBTM116-D9

Size: 8176 bp

Promoter/terminator: Truncated ADH

DNA-binding domain: LexA (N-terminal)

Selection marker: Tryptophan (TRP1)

Antibiotic marker: Tetracycline

Sequencing primers: BTM-5plus, BTM-3min

Properties Y2H vector: 2µ plasmid

Reference: Goehler *et al.,* 2004

pBTMcC24-DM

Size: 10871 bp

Promoter/terminator: Truncated ADH

DNA-binding domain: LexA (C-terminal)

Selection marker: Tryptophan (TRP1)

Antibiotic marker: Tetracycline

Sequencing primers: BTM-5plus, BTM-3min

Properties Y2H vector: 2µ plasmid

Reference: Stelzl laboratory (not published)

pACT4-DM

Size: 9613 bp

Promoter/terminator: Truncated ADH

Activation domain: GAL4 (N-terminal)

Selection marker: Leucine (LEU2)

Antibiotic marker: Ampicillin

Sequencing primers: Prey-5p

Properties Y2H vector: 2µ plasmid

Reference: Goehler *et al.,* 2004

2.7.4 Expression vectors for mammalian cell lines

pFireV5-DM

Size: 8828 bp

Promoter: Human cytomegalovirus (CMV) immediumte-early promoter

Terminator: Bovine growth hormone (BGH) polyadenylation signal

V5-firefly (N-terminal)

Antibiotic marker: Ampicillin

Sequencing primers: T7-for/BGHrev

Reference: modified from Palidwor *et al.*, 2009

pcDNA3.1PA-D57

Size: 7646 bp

Promoter: Human cytomegalovirus (CMV) immediumte-early promoter

Terminator: Bovine growth hormone (BGH) polyadenylation signal

Protein A (N-terminal)

Antibiotic marker: Ampicillin

Sequencing primers: T7-for/BGHrev

Reference: modified from Palidwor *et al., 2009*

2.8 Oligonucleotides

uni_attB1 for	GGGGACAAGTTTGTACAAAAAAGCAGGCT
uni_attB2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGT
PRMT3_attB1_for_domain	AAAAAGCAGGCTTAATGGGGCATTATGGGATACATGAA
PRMT3_attB2_rev_domain	AGAAAGCTGGGTCCTACAAATCTGAGATAGACGTCGTA
PRMT3_attB1_for_fl	AAAAAGCAGGCTTAATGTGCTCGTTAGCGTCAGG
PRMT3_attB2_rev_fl	AGAAAGCTGGGTCTCACTGGAGACCATAAGTTTG
PRMT6_attB1_for	AAAAAGCAGGCTTAATGATCGCGGACCGCGTCCG
PRMT6_attB2_rev	AGAAAGCTGGGTCCTAGTCCTCCATGGCAAAGT
SETD7_attB1_for_domain	AAAAAGCAGGCTTAATGGATAGCGACGATGAGGT
SETD7_attB2_rev_domain	AGAAAGCTGGGTCTCACTTCTGCTGGGTGGCC
pACT_5p:	CCAAAGCTTCTGAATAAGCC
pACT_3p:	AATTAATTCCCGAGCCTCCA

2.9 Antibodies

Antibody: Sheep gamma globulin, host species: sheep, Dilution: 1:1000, Source Jackson ImmunoResearch (013-000-002)

Antibody: AffiniPure rabbit anti sheep IgG, host species: rabbit, Dilution: 1:750, source: Jackson ImmunoResearch (313-005-003)

2.10 Databases

Swissprot http://www.ebi.ac.uk/swissprot/

Expasy http://us.expasy.org/tools/

Interpro http://www.ebi.ac.uk/interpro/

Gene ontology http://www.geneontology.org/

National Center Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/

2.11 Software and tools

Access

ChemSketch http://www.acdlabs.com/download/

Clustalw http://www.ebi.ac.uk/Tools/msa/clustalw2/

Cytoscape http://www.cytoscape.org/

DAVID http://david.abcc.ncifcrf.gov/

Itol http://itol.embl.de/

MaxQuant http://maxquant.org/

VectorNTI Invitrogen

VisualGrid

3 Methods

3.1 Microbiology and molecular biology

3.1.1 Growth and storage of E. coli

Liquid *E. coli* cultures were generated by inoculation of LB growth medium containing appropriate antibiotics with a single *E. coli* colony. Cells were grown in a shaker at 37 °C for 16-20 hours. Glycerol stocks allow long term storage of *E. coli* and were generated by mixing an overnight culture with autoclaved 50 % glycerol at a ratio 1:1 and were stored at -80 °C. For short term storage *E. coli* strains were streaked out on LB agar plates containing the appropriate antibiotics and incubated at 37 °C for 16-20 hours. The agar plates with grown *E. coli* colonies were rapped in foil and stored at 4 °C for 1 month.

3.1.2 Preparation of competent E. coli cells

3.1.2.1 Chemically competent cells

To generate chemical competent cells 2l 2YT medium was inoculated with an overnight culture to an OD_{600} less than 0.1. Cells were grown at 37 °C while shaking until an OD_{600} of 0.7-0.8 was reached. The culture was centrifuged at 4 °C at 1258 x g. After removing the supernatant the pellet was resuspended in 40 ml sterile TSS. Four milliliter 87 % glycerol was added to yield the final competent cell suspension which was aliquoted, frozen in liquid nitrogen and stored at -80 °C.

3.1.2.2 Electrocompetent cells

To generate electrocompetent competent cells 1.5 l LB growth medium were inoculated with an overnight culture to an OD_{600} less than 0.1. Cells were grown at 37 °C while shaking until an OD_{600} of 0.4-0.6 was reached. The culture was incubated on ice for 30 minutes and then centrifuged (1258 x g) at 4 °C for 30 minutes. After removing the supernatant the pellet was resuspended and washed with 1 l cold water. After centrifugation and removing the supernatant the pellet was resuspended in 600 ml cold and sterile 15 % glycerol. The suspension was pelleted again and resuspended in 12 ml of cold 10 % glycerol to yield the final competent cell suspension which was aliquoted, frozen in liquid nitrogen and stored at -80 °C.

3.1.3 Transformation of competent E. coli cells

3.1.3.1 Chemical transformation of competent *cells*

The protocol describes a 96-well format transformation of *E. coli* cells with the BP or LR reaction mixture. Supplement 3 µl of the reaction mixture to a 96-well PCR plate. Competent bacterial cells

were thawed on ice and 30 µl pipetted on top of the BP or LR reaction. The PCR plate was sealed with plastic tape, vortexed softly and incubated on ice for 30 minutes. Then the plate was incubated at 42 °C for 90 seconds and returned to ice for further 5 minutes. To the reaction mixture 70 µl from 200 µl prewarmed SOC medium were added. Then the whole content of the PCR plate wells was transferred into the corresponding wells of a deepwell plate, containing the remaining 130 µl prewarmed SOC medium. The deepwell plate was shacked at 37 °C for 1 hour. Supplement 50 µl from each well onto selective LB plates by making rows of drops on the agar. LB plates were incubated at 37 °C for 16-20 hours. This protocol yields colonies for 95-100 % of the BP or LR reactions if chemically competent DH10B cells were used. If no colonies were obtained transformation was repeated with 1-2 µl of the remaining LR reaction using the electroporation protocol.

3.1.3.2 Electroporation of competent *E. coli*

Typically for an electroporation of BP or LR reaction 1 µl of the reaction solution was sufficient and pipetted into 30 µl thawed electrocompetent cells. The mixture of cells and DNA was transferred into a 0.1 cm gap electroporation cuvette and tapped to the bottom. An electroporator was used to pulse the sample in the cuvette, generating one pulse with field strength of 1.8 kV and a time constant of approximately 5 milliseconds. Immediately after pulse delivery 200 µl of the 500 µl prewarmed SOC-Medium was added. The cell suspension was transferred into a sterile 2 ml microfuge tube containing the remaining 300 µl medium and shaked at 37 °C for 1 hour. Supplement 200 µl cell suspension on LB agar containing the appropriate antibiotics and incubated overnight at 37 °C.

3.1.4 Plasmid isolation of E. coli

This 96-mini-prep protocol produces DNA used for LR reactions, sequencing and yeast transformations. Deep well plates containing 1.2 ml LB medium per well were inoculated, sealed with a breathable sealing film and incubated on a shaker (Heydorf, 1000 rpm) at 37 °C. After 16-20 hours glycerol stocks were created. The remaining cells were collected by centrifugation (1258 x g) at 4 °C for 30 minutes. After removing the supernatant the pellets were resuspended in 300 µl of cold Buffer P1 (containing RNase A) by vigorous vortexing for 2-3 minutes. Next, 300 µl Buffer P2 were added to each well and the resealed plates were mixed thoroughly by inverting the plates 3-4 times. After 5 minutes of incubation at room temperature 300 µl Buffer P3 were added to each well and the resealed plates were mixed thoroughly as described above. The lysate was cleared by centrifugation (3220 x g) for 1 hour and 750 µl were transferred into the corresponding wells of a new deepwell plate. The plate was sealed after adding 530 µl isopropanol. Then it was mixed thoroughly and centrifuged (3220 x g) at room temperature for 1 hour in order to precipitate the plasmid DNA. The supernatant was omitted and 1 ml 70 % ethanol was added to each well. The plate was sealed and centrifuged (3220 x g) at 4 °C for 1 hour. After removing the supernatant the

pellets were air dried for 30-60 minutes and dissolved in 100 µl sterile water. The success of the mini-prep was analyzed by BsrGI restriction analysis.

Commercially available mini-prep kits (Qiagen) and midi-prep kits (Promega) were undertaken according to the standard protocol in the manufactures guidelines.

3.1.5 Restriction digest of plasmid DNA

The restriction endonuclease FD-Bsp1407I, a fast digest isoschizomer of BsrGI, was used to control the success of mini- and midi-preps by restriction analysis. Typically 400 ng DNA sample were mixed with restriction endonuclease mix (1x FD-Buffer, 0.1 µl FD-Bsp1407I) and incubated at 37 °C for 30 minutes.

3.1.6 Separation of DNA fragments by agarose gel electrophoresis

Agarose gel electrophoresis was used to control the plasmid insert size after restriction enzyme digest. Agarose was dissolved in 0.5 x TBE buffer to a final concentration of 1 %. DNA samples of 200 ng were supplemented with Orange G Sample buffer and loaded onto agarose gels alongside 4 µl of a 1 Kb Plus DNA ladder for band size estimation. After electrophoresis at 130 V for 1 hour using 0.5x TBE as running buffer, the gels were stained for 15 minutes in a 1:20,000 dilution of SYBR Gold nucleic acid gel stain in 0.5x TBE. DNA bands were visualized using an ultraviolet light source.

3.1.7 Determination of DNA concentration

The concentration of DNA in aqueous solution was determined by measuring the absorbance at 260 nm in a spectrophotometer (Nanodrop 2000). Nuclease free water was used as standard and the absorbance of 1 µl DNA solution pipetted onto the Nanodrop was measured.

3.1.8 Polymerase chain reaction (PCR)

In order to generate gateway compatible clones of PRMT3, PRMT6 and SETD7 the two step gateway PCR followed by the BP reaction was performed. The full length or domain of the PMT was amplified and gateway recombination sites were added in two consecutive amplification steps. Gene-specific PCR primers were used to amplify the ORF in the first round of PCR, adding minimal overhangs that contain part of the gateway attB-sequences. These overhangs were AAAAAGCAGGCTTA followed by the gene specific bases for the forward primer and AGAAAGCTGGGTC followed by the gene specific bases for the reverse primer. The second PCR was carried out with two “universal gateway primers” that extend the attB-sequences to generate complete recombination sites. In the first amplification step the 20 µl PCR reaction mix contain 2 ng entry vector, 0.5 µM gene-specific primers, 0.02 U/µl Phusion hot start, 0.2 mM dNTP mix, 3 % DMSO and 1x Phusion reaction buffers. The reaction volume was adjusted to 20 µl with

sterile water before the reaction mixture was cycled in a PCR thermal cycler. The PCR was initialized by heating the reaction to a temperature of 98 °C for 30 seconds before ten cycles of denaturation (98 °C for 8 seconds), annealing (63 °C for 23 seconds) and extension (72 °C for 45 seconds) were performed. After a final extension step at 72 °C for 8 minutes the reaction was cooled down to 4 °C. This first PCR product was used as template in a second PCR which was performed with "universal gateway primers". For the second PCR reaction 5 µl of the obtained PCR product were transferred into a new PCR tube to which the components were added as before, except the primers were exchanged by the universal primers. The PCR was initialized by heating the reaction to a temperature of 98 °C for 30 seconds, then five cycles of denaturation (98 °C for 8 seconds), annealing (45 °C for 30 seconds) and extension (72 °C for 60 seconds) were performed. After further twenty cycles with increased annealing temperature (60 °C) and a final extension step (72 °C for 8 minutes) the reaction was cooled down to 4 °C. The PCR product was analyzed using agarose gel electrophoresis to ascertain the correct band size and estimation the DNA concentration prior PCR purification. Commercially available PCR Purification Kit (Qiagen) was used according to the standard protocol in the manufactures guideline.

3.1.9 Gateway cloning technology

3.1.9.1 BP Reaction

Purified PCR products obtained from the two step PCR reaction were combined with the donor vector pDONR221 to generate an entry vector. The BP reaction set up on ice was composed of 75 ng entry vector, 20 ng PCR product and 1 µl BP clonase enzyme mix. The reaction was mixed and incubated at 25 °C for 3-18 hours. The reaction were used directly in a bacterial transformation or stored at -20 °C.

3.1.9.2 LR Reaction

In order to rapidly transfer ORFs into expression vectors for further study a LR reaction was undertaken followed by transformation into bacterial cells. The LR reaction was setup on ice and each reaction was composed of 75 ng destination vector, 200-300 ng entry vector and 1 µl LR clonase enzyme mix. The reaction was incubated at 25 °C for 3-18 hours to allow site specific recombination. The reaction were directly used for transformation or stored at -20 °C.

3.1.10 Expression of recombinant fusion proteins

Candidate proteins were expressed as recombinant fusion in *E. coli*, purified from the cell lysate and analysed in a methylation assay. The SCS1 *E. coli* strain carrying the helper plasmid pRARE was transformed with the hexahistidine-tag (his-tag) expression vector pDESTco. The Rosetta strain from Novagen contains the pRARE plasmid that expresses rare tRNAs. 2YT medium supplemented with appropriate antibiotics and 2 % glucose were inoculated with the desired

expression clone from a frozen stock and incubated at 37º C in a shaker for 16-20 hours. Subsequently 100 ml prewarm SB medium containing appropriate antibiotics and 20 µg/ml thiamine were inoculated with the overnight culture to an OD_{600} less than 0.1. Incubation was continued until an OD_{600} of 0.7-0.8 was reached. For induction of protein expression IPTG was added to a final concentration of 1 mM and incubation was continued at 30 °C for further 4 hours. Cells were harvested by centrifugation at 1258 x g for 20 minutes. The cell pellets were stored at -80 ºC and later on used to purify his-tag proteins over Ni-NTA agarose beads. Expression of his-tagged protein was controlled on SDS-PAGE or by western blotting.

3.2 Yeast specific molecular biology

3.2.1 Preparation of yeast media

This section describes the preparation of different media from stock solutions. The media were named after the missing and required amino acids or nucleosides. Amino acids or nucleosides were abbreviated with a single letter as followed: “H” for histidine, “A” for adenine, “U” for uracil, “L” for leucin and “T” for tryptophan.

Liquid medium: 25 ml 20x glucose stock solution and 5 ml of each required 100x amino acid or nucleoside stock solution were added to 400 ml 1.25x NB or 1.25x NBG. Liquid medium was adjusted to a final volume of 500 ml with sterile water.

Solid medium: 200 ml 2.5x NB, 25 ml 20x glucose stock solution and 5 ml of each required 100x amino acid or nucleoside stock solution were added to 200 ml 2.5x agar. Solid medium was adjusted to a final volume of 500 ml with sterile water and dissolved using a microwave. After the medium was cooled down to 60 °C agar plates were filled with 200 ml medium under a sterile hood.

YPD liquid medium: 25 ml 20x glucose stock solution were added to 400 ml 1.25x YPD. YPD liquid medium was adjusted to a final volume of 500 ml with sterile water.

YPD solid medium: 25 ml 20x glucose stock solution were added to 400 ml 1.25x YPD agar. Solid medium was adjusted to a final volume of 500 ml with sterile water and dissolved using a microwave. After the medium was cooled down to 60 °C agar plates were filled with 200 ml medium under a sterile hood.

3.2.2 Growth and storage of S. cerevisiae

Correct growth and storage of yeast is important for successful screening. To generate liquid cultures of yeast the requested amount of liquid medium was inoculated with yeast colonies and vortexed. Cells were grown in a shaker (Innova44, 250 rpm) at 30 °C for 16-20 hours. Glycerol stocks enable long-term storage of yeast and were generated by freezing (-80 °C) overnight cultures grown in NBG medium. These glycerol stocks can be thawed twice. For short-term storage yeast strains were streaked on solid medium and incubated at 30 °C for 1-5 days prior to storage at 4 °C.

Stored yeast was replicated on agar for an extra generation before starting a Y2H experiment. Generally, untransformed yeast was grown in YPD medium and transformed yeast in the appropriate selective medium.

3.2.3 Transformation of S. cerevisiae

The MATa yeast strain was transformed with different bait plasmids. The following protocol describes the transformation of 96-well plates. Fresh MATa yeast were inoculated in YPD liquid medium, vortexed and grown in a shaking incubator (Innova44, 250 rpm) at 30 °C for 16-20 hours. The YPD medium was inoculated with the overnight culture to an OD_{600} of 0.10-0.15 and then incubated at 30 °C until an OD_{600} of 0.6 - 0.8 was reached. At this point the culture was harvested by centrifugation (1258 x g) at room temperature for 5 minutes. After removing the supernatant each pellet was resuspended in sterile 1x TE and then again pelleted as above. After removing the supernatant the pellet was resuspended in 2000 µl Mix I, pooled and incubated at room temperature for 10-120 minutes. The yeast suspension was mixed well before transferring 11 µl into each well of a PCR plate containing a mixture of 2.5 µl bait plasmid DNA and 5 µl 10.5 mg/ml heat denatured salmon testis DNA, serving as carrier DNA. One negative control (*i.e.* only carrier DNA) and one positive control (*i.e.* a well-tried vector preparation) were included in each plate. The yeast suspension was mixed with the DNA. Into each well of the PCR plate 58 µl of Mix 2 were added. The plate was sealed, mixed and incubated at 30 °C for 30 minutes before 8 µl DMSO were added into each well. The plate was sealed, mixed and incubated at 42 °C in a thermocycler for 7 minutes. Four biological replicas were generated by transferring the cells to four selective agar plates. Selection of the transformed yeast cells required -T/HUL medium because the bait vector (pBTM116-D9 or pBTMcC24) has a tryptophan selection marker. Plates were incubated at 30 °C for 3-5 days to allow colonies to grow and indicate positive transformation. Colonies obtained from the transformation were replicated before a glycerol stock was prepared, and they were used in an auto-activation or mating experiment.

3.2.4 Autoactivation test of baits

Autoactive baits induce constitutive activation of reporter genes even in the absence of a positive interacting protein conjugated to the Gal4 domain. This causes false positive interactions in Y2H screens and masks any interaction signal of other baits in a pooled Y2H approach. In order to identify autoactive bait strains the four biological replicas were mated with a prey strain carrying a prey plasmid without insert. Baits that were autoactive grew on -HULT medium and were not taken forward into the pooled Y2H-matrix approach. One day prior mating liquid -L/HAUT medium was inoculated with freshly grown prey strain carrying a prey plasmid without insert. After vortexing and incubation at 30 °C in a shaker (Innova44, 250 rpm) for 16-20 hours 100 µl liquid culture were pipetted into each well of a 96-well MTP. Using a pin tool bait strains freshly grown on selective agar were stirred into the MTPs containing the prey strain without insert. This bait and prey strain

mixture was directly stamped onto YPD agar and incubated at 30 °C for 36-44 hours. Yeast was resuspended in -LT/HU filled MTPs and transferred to -LT/HU agar to select for diploid yeast. After 4 days at 30 °C the yeast was transferred to -HULT agar (via -LT/HU filled MTPs). The colonies were left to grow at 30 °C for 5-7 days to select for growth reporter gene activity. Bait strains which grew on -HULT agar plates and were not taken forward into pool screening approaches.

3.2.5 Screen of bait pools against a prey array (Y2H-matrix approach)

Prey and bait strains were replicated in 384- or 96-well format, respectively. Three to four days before they were used in the Y2H-matrix mating experiment. One day before mating, bait strains freshly grown on selective agar were stirred into 96-well MTPs containing 100 µl -T/HUL liquid medium per well. In the next step 10 µl of this yeast suspension was used to inoculate deepwell plates containing 1.2 ml -T/HUL medium per well. The baits were grown separately to the early stationary phase (OD_{600} =1.5-3) at 30 °C in a shaker for 18-22 hours. The four bait replicas were seperatly pooled and kept separately during the screen. Into each well of a 384-well MTP 38 µl of the pooled overnight culture were pipetted. Using a pin tool prey strains freshly grown on selective agar were stirred into the MTPs containing the bait pools. This bait and prey strain mixture was directly stamped onto YPD agar and incubated at 30 °C for 36-44 hours. Yeast was resuspended in -LT/HU filled MTPs and transferred to -HULT agar. To control for efficient mating some were also stamped on -LT/HU agar. The plates were incubated at 30 °C for 5-7 days to allow selective growth indicating positive interaction between a specific prey and at least one bait in the pool (-HULT agar). The interacting preys were identified over the prey array position.

3.2.6 Screen of bait pool against a pooled prey array (Y2H-seq approach)

The prey array and selected bait strains were replicated (in 384- or 96-well format, respectively) 3-4 days before they were used in the Y2H-seq mating experiment. The fresh grown yeast was scratched of the agar with a sterile spittle. The prey array was resuspended in 40 ml and the four bait replicas each in 10 ml of the corresponding liquid media. In a dilution of 1:200 an OD_{600} of 1.6 was measured for the pooled prey array and OD_{600} of around 0.8 were measured for each of the four bait replicas. Ten ml of each bait pool were mixed with 5 ml of the pooled prey array to give equal amounts of bait and prey. The pooled and mixed cultures were distributed into one 384-well MTP and gridded repeatedly onto YPD agar. Mating took place in 13,824 spots at 30 °C for 36-44 hours. Yeast colonies were resuspended in 36 384-well micro titer plates containing -LT/HU liquid medium and gridded in high density 9x384 format onto -HULT agar plates for interaction selection (124,416 spots). To control for efficient mating some were also stamped on -LT/HU agar. The plates were incubated at 30 °C for 5-7 days to allow growth on selective media indicating positive interaction between at least one pair of bait and prey. All yeast colonies grown on –HULT agar were scratched off and the pellet stored at -80 °C.

3.2.6.1 Preparation of plasmid DNA

From the yeast cells grown on selective medium plasmid DNA was isolated. Therefore, the pelleted yeast cells were resuspended in 7 ml H1 buffer (containing lysozyme and RNase), inverted 3-4 times and incubate at 37 °C for 45-60 minutes. To lyse the cells 7 ml P2 was added. To the viscous turned suspension 7 ml P3 were added and after incubation for 1 hour the suspension was centrifuged (3220 x g) for 90 minutes. The supernatant was pipetted into a new tube. The DNA was precipitated with 0.7 [v/v] 2-propanol, centrifuged for 1 hour at 3220 x g and the supernatant discard. The pellet was dried and resuspend in 1 ml 1x TE.

The sample was purified by phenol extraction. Therefore, 0.5 [v/v] phenol and 0.5 [v/v] chloroform (isoamylalkohol 24:1) were added. The suspension was vortexed, centrifuged and then the upper liquid phase pipetted into a new tube. Again, phenol and chloroform were added and steps were performed as described above. To remove the remaining phenol 1 [v/v] chloroform was added. The suspension was vortexed, centrifuged and the upper liquid phase pipetted into a new tube.

The samples were dialyzed to get ride of small size contaminants like RNAs. Therefore, the sample was pipetted into a dialyze pipe and incubated for one hours in 5 liter baker glass filled with 1x TE. This step was repeated three times with incubation time of two hours, three hours and 16 hours at 4 °C. To decrease the enlarged volume of the samples the DNA was precipitated by adding 0.1 [v/v] 3 M NaOAc (pH 6.0) and 3 [v/v] 99 % ethanol. Then the suspension was mixed and incubated at -20 °C for 20 minutes. Centrifugation (3220 x g) revealed a visible pellet. After removing the supernatant the pellet was dried and dissolved in sterile water. The plasmid DNA was controlled on an agarose gel.

3.2.6.2 Amplification of prey plasmid DNA

Plasmid DNA was isolated from the yeast and the inserts, coding for the proteins of interest, were PCR amplified with 5'-hex-blocked primer (the 3' nucleoside was attached via a phosphothioate) that were targeted against unique sites 5' and 3' to the ORF insert in the prey plasmid (pACT_3p and pACT_5p). The prepared plasmid DNA from the selective grown yeast was diluted 1:10 and used in a PCR reaction. The reaction mix contains 10 µl plasmid DNA (1:10), 1x HF Phusion buffer, 0.2 mM dNTPs mix, 1 µM primer, 3 % DMSO and the 0.02 U/µl Phusion polymerase. The reaction volume was adjusted to 50 µl with sterile water before the reaction mixture was heated up to 98 °C for 30 seconds to activate Phusion polymerase. Twenty-five cycles were performed starting with 98 °C for 8 seconds, annealing temperature of 55 °C for 23 seconds and elongation temperature of 72 °C for 60 seconds, plus a final extension step of 72 °C for 8 minutes. The PCR reaction was controlled by agarose gel electrophoresis to ascertain plasmids DNA of variable length were amplified. Hence, the PCR yielded a DNA smear of the expected size range (800 to > 4000 bp) for the prey ORFs when applied to agarose gel electrophoresis. Twenty-three PCR

reactions were combined and purified by the QIAquick PCR purification Kit according to the standard protocol in the manufactures guideline. The total amount of the purified PCR product was estimated to be around 5 µg by comparison to known concentrated DNA probes on an agarose gel. Minimal 3 µg of total DNA were needed for second generation sequencing. The amplified prey plasmid DNA was further amplified, fragmented and sequenced using the Illumina Genome Analyzer (by the second generation sequencing team, Bernd Timmermann, MPIMG in Berlin).

3.2.6.3 Computational analysis of second generation sequencing readout

High quality 36 base sequences were extracted from the raw sequence data (sequences containing no undetermined base calls). This resulted in ~34-49 million reads per sequencing run distributed over ~6-9 million unique sequences, highlighting that each unique 36 base read can be represented multiple times. These sequences were then mapped against a pseudo genome composed of the human protein coding sequences from RefSeq (NCBI release 2011-02-16), the *S. cerevisiae* genome and the Y2H vector sequences. To map the HQ sequences against the pseudo genome the SHRiMP software package, version 1.2.1, was utilized with default settings (David et al., 2011). Only reads with perfect alignment to target sequences were taken forward for further analysis.
In order to generate a ranked retest list of Y2H prey clones that accurately reflects the mappings, a sequence score was calculated for every RefSeq identifier. To bias against RefSeq identifiers that were mapped with a high number of mappings but with a low number of unique reads, a sequence score (SeqScore) combining the number of mappings and the number of unique reads for each RefSeq protein coding sequence (in both forward and reverse direction) was utilized.

$$\text{SeqScore} = \frac{(cds - 35) * 2}{\sum_{k=1}^{n} \frac{1}{x_k} + \left((cds - 35) * 2 - n\right)}$$

cds is the protein coding sequence length (bps); *x* is the number of mappings associated with each unique read +1; *n* is the number of unique reads matching a given Refseq coding sequence. The SeqScore gives a relatively high value for RefSeq identifiers that were mapped with multiple unique sequences and with high frequency as potential interacting preys for retesting. Entrez GeneID based ranked retest lists were generated by taking the maximum SeqScore from the set of associated Refseq identifiers for each analytical workflow. For the final retest list we took the second highest score for each gene ID from the four ranked lists. We therefore excluded gene IDs associated with a high ranking score from only one biological replicate favoring those that scored highly over at least two. This was done in cooperation with Jonathan Woodsmith and Arndt Grossman (OWL Stelzl, MPIMG in Berlin).

3.2.7 Retest of primary hits

To verify primary hits obtained in the Y2H-matrix and Y2H-seq screen a retest was performed. The retest is identical for both approaches. Non-auto active preys that grew in more than one replica of a pool were included in the retest and were picked from the prey array. The four replicas of the baits were combined in one 384-well MTPs. Prey and bait strains were replicated 3-4 days before they were used in the retest mating experiment. One day before mating, fresh grown yeast spots containing the prey protein were stirred into 20-40 ml -L/HUT liquid medium using an inoculation loop. After vortexing and incubation at 30 °C in a shaker for 16-20 hours, liquid cultures were pipetted into the wells of one 384-well MTP. Using a pin tool bait strain replicas freshly grown on selective agar were stirred into the MTPs containing the preys. This bait and prey strain mixture was directly stamped onto YPD agar and incubated for 36-44 hours at 30 °C. Yeast was resuspended in -LT/HU filled MTPs and transferred to -LT/HU and -HULT agar. The plates were incubated at 30 °C for 5-7 days to allow selective growth indicating successful mating (-ALT/HU agar) or a positive interaction between a specific bait-prey pair (-HAULT agar).

For more detailed description of yeast two-hybrid screening approach see Worseck et al (2012).

3.3 Mammalian cell culture

3.3.1 Cell culture and transfection

HEK 293T were cultured in a humidified 5 % CO_2 atmosphere at 37 °C in Dulbecco's Modified Eagle's Medium (DMEM) with 10 % fetal bovine serum. Cells were maintained by growing to ~80 % confluence in 75 cm flasks prior trypsinization and splitting at a 1:10 dilution every 3-4 days. Cells were seeded into a 10 cm petri dish. At the time 80 % confluence was reached the cells were transfected. Cells were transiently transfected with 640 ng per petri dish using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. The cells were incubated at 37 °C in a CO_2 incubator for 20-42 hours.

3.3.2 Preparation of cell lysate for western blots

HEK293 cells were seeded in 96well plates and transfected as described above. Two days after transfection, the DMEM medium was removed and cells were lysed by addition of 20 µl 4x SDS gel-loading buffer per well. Samples originating from triplicate transfections were pooled in one tube and heated to 95 °C for 5 minutes prior storage at -20 °C or analysis by SDS-PAGE.

3.4 Protein biochemistry

3.4.1 SDS-polyacrylamide gel electrophoresis

In SDS-polyacrylamide gel electrophoresis (SDS-PAGE) proteins were separated mainly on the

basis of polypeptide length. Described below is the protocol for preparing and using Laemmli discontinuous gels. This system uses two sequential gels. The top gel, called the stacking gel, was slightly acidic (pH 6.8) and has a low (5 %) acrylamide concentration to make a porous gel. The lower gel, called the separating or resolving gel, was more basic (pH 8.8) and has acrylamide concentrations of 12 %. Gels were poured and assembled in a Mini-Protean electrophoresis system (Bio-Rad). For 20 ml resolving gel solution; 5 ml 1.5 M Tris pH 8.8, 200 µl 10 % SDS, 150 µl 10 % APS and 15 µl TEMED were supplemented with 5 ml (10 %), 6 ml (12 %) or 7 ml (14 %) Rotiphorese Gel 40 (37.5:1) and adjusted to a final volume of 20 ml with water. Between glass plates 7 ml of this resolving gel solution were poured and isopropanol was placed above the gel to provide a smooth surface. After the gel was completely polymerized, the isopropanol was poured off and a standard 5 % acrylamide stacking gel was added (for a 10 ml solution: 1.25 ml Rotiphorese Gel 40 (37.5:1), 1.25 ml 1 M Tris pH 6.8, 7.3 ml H2O, 100 µl 10 % SDS, 100 µl 10 % APS and 20 µl TEMED) and the comb placed into position at the top of the glass plates. Heat denatured protein samples in SDS gel loading buffer were loaded alongside with 5 µl of a prestained protein ladder (PageRuler™ Plus, Fermentas) for band size estimation. Typically gels were run in 1x electrophoresis running buffer (25 mM Tris, 0.1 % SDS and 192 mM glycine) at 90 V for 15 minutes followed by 140 V for 1 hour. Gels were used for analysis of specific proteins through western blotting and estimation of protein content through Coomassie blue or blue silver staining.

3.4.2 Protein gel stain

3.4.2.1 Coomassie blue

The gel was soaked in the Coomassie blue stain (50 % (v/v) methanol, 10 % (v/v) acetic acid, 2.5 % (w/v) Coomassie Brilliant Blue G-250) on a rocker for 1 hour at room temperature. Excess dye was eluted with detain solution (30 % (v/v) methanol, 10 % (v/v) acetic acid). Stained gels were transferred to a light table and images were taken.

3.4.2.2 Blue silver

The gels were soaked in blue silver stain (20 % (v/v) methanol, 10 % (v/v) phosphoric acid, 10 % (w/v) ammonium sulfate, 0.12 % (w/v) Coomassie Brilliant Blue G-250) until protein band were visible and excess dye was eluted with water. Stained gels were transferred to a light table and images were taken.

3.4.3 Western blot

To detect specific proteins via antibody proteins from the SDS gels were transferred to a nitrocellulose membrane using the semi-dry transfer. Six 3 mm thick Whatman paper and one PVDF membrane at the size of the gel were incubated in transfer buffer. Three sheets of Whatman

paper, the nitrocellulose membrane, the gel and again three sheets of Whatman paper were put on top of each other to transfer the proteins to the PVDF membrane. Air bubbles between the layers need to be removed. The upper (negative) electrode plate was put on top of the stack and a constant current of 55 mA per gel was applied for 1 hour. After that, the membranes were incubated in 5 % BSA in TBST (blocking buffer) for at least 1 hour at room temperature on a rocker to block non-specific binding of antibodies. Membranes were then incubated at room temperature for 1 hour or at 4 °C overnight on a rocker with a specific primary antibody diluted in blocking buffer. Afterwards, membranes were washed three times in TBST for 10 minutes prior to incubation with a secondary antibody diluted in blocking buffer for 1 hour. Membranes were then washed three times in TBST for 10 minutes and afterwards the antibodies bound to the specific proteins were visualization by immunoreactive bands using a high resolution CCD camera (e.g. Fuji LAS 3000) according to the manufacturer's instructions. Depending on the secondary antibody the bands were visualized by adding 1 ml of the luminol reagent ECL (Enhanced Chemiluminescence Substrate) or the fluorimetric substrate AttoPhos. In contrast to the chemiluminescence detection method (ECL) the chemifluorescence detection method (AttoPhos) depends on illumination with blue LED light (460 nm). Additionally, for chemifluorescence signal detection a Y515-D emission filter for blue LED light was screwed on the CCD camera. The dilutions used for each primary and secondary antibody are listed above.

3.4.4 His-tag protein purification

E. coli pellet of 100 ml culture was resuspended in 500 µl NiNTA resuspension buffer and sonificated on ice. Lysate was centrifuged (17,949 x g) at 4 °C for 10 minutes. The supernatant transferred into a new tube containing 100 µl equilibrated Ni-NTA agarose beads. Probes were incubated at 10 °C on a rotator for 1 hour. The suspension was centrifuged for 30 seconds at 106 x g and supernatant removed. Beads were washed with 800 µl wash buffer, centrifuged and supernatant removed. His-tagged proteins bound to the beads were used in methylation assay as substrate or were eluted with SDS loading buffer to control for protein expression on a SDS-PAGE.

3.4.5 Methylation assay

To detect methylation the expressed and purified candidate his-fusion proteins were incubated with mammalian cell lysate overexpressing the protein methyltransferase of interested. Therefore, petri dishes with PRMT1 or CARM1 transfected HEK293 cells were washed with cold PBS. Cells were dissolved from the culture plate with a plastic scraper, transferred into a 15 ml falcon and sonificated on ice. Cell debris was removed by centrifugation and the supernatant used in the methylation assay. 100 µl supernatant were mixed with 5 µl ^{3}H AdoMet (2.75 µCIE from a 0.55 CIE/ml stock solution) and incubated with 1 µg of his-fusion protein immobilized on Ni-NTA agarose. After 2 hours incubation at 30 °C in a thermo mixer (750 rpm) the immobilized proteins were washed five times with cold wash buffer. After that, the beads were transferred into a vial,

1 ml scintillation liquid was added and samples were rocked at room temperature for 16-20 hours. The ionizing radiation of the tritium was measure in a scintillation counter for 5 minutes.

In a non radioactive methylation assay 100 µM non labeled AdoMet (^{1}H AdoMet) was used instead of ^{3}H AdoMet. After incubation as described above the beads were washed and SDS gel-loading buffer was added. The samples were separated on an SDS-PAGE and the band corresponding to the protein of interest was cut out. The gel slice was alkylated with iodoacetamide, trypsin digested and subjected to LC-MS/MS analysis (by David Meierhofer, OWL Sauer, MPIMG in Berlin).

Data were analyzed by MaxQuant (http://maxquant.org/) quantitative proteomics software package to analyze large mass-spectrometric datasets. We scanned for mono- and dimethylarginines as well as for mono-, di- and trimethyllysines. Spectra of methylated and unmethylated peptides were manually inspected.

3.5 Protein interaction analysis

3.5.1 Database to store the data

If dealing with HTP data it was crucial to minimize error through accurate and high-quality data storage. All generated entry or destination vectors alongside with all tested Y2H interactions were stored in a relational database (SQL). To remove duplications or clarity issues, CloneIDs were used as unique identifier for each ORF. The CloneID was stored together with the clone position (bank, plate, row and column) and additional data (results of BsrGI restriction analysis and backbone vector). All tested Y2H interactions were recorded together with the corresponding screen-, agar- and plate-number including the plate-position as identifier and the final interaction result. This simplifies the analysis of the obtained data.

3.5.2 R statistical software to generate a heatmap

To visualize the specificity of bait interactions a heatmap was generated. Therefore, the number of preys interacting with the enzyme of interessed and another enzyme was divided by the total number of preys interacting with the enzyme. To generate a heatmap the gplots software package, version 2.10.0, was utilized with default settings was used.

3.5.3 Perl script to calculate RG-repeat score

The RG-repeat score was calculated with a perl script under UNIX. In a window of 12 amino acids RG-repeats were counted. The score was calculated by multiplying the count of repeats with the count of repeats and this was divided by times found a window containing repeats.

3.5.4 Cytoscape to visualize obtained PPI data

Cytoscape (version 2.7.0) was used for visualization of the obtained PPI data in form of graphs. In the interaction networks nodes represent proteins, while edges connecting these nodes represent obtained interactions between these proteins.

3.5.5 DAVID to discover enriched functional-related gene groups

Enrichment analysis was performed with DAVID (Huang et al., 2009) using 324 identified preys (Entrez GeneID level) with the Y2H array genes as background. The tool reports the modified Fisher exact p-value for enrichment and the absolute number of genes for each annotation cluster. The domain annotation, the cellular component and the biological function of the proteins were downloaded from the website and integrated in the cytoscape network.

4 Results

4.1 Y2H screening as PPIs discovery tool

The Y2H-matrix screen approach is one of the most efficient methods to discover proteome-wide protein-protein interactions (PPIs) (Worseck et al., 2012). It produces high-quality, binary PPI data and is specifically suitable for detection of transient interactions. However, it provides a low sensitivity (Vinayagam et al., 2010; Yu et al., 2008). The sensitivity of individual Y2H systems is only 5 to 20 %, which is comparable to other PPI detection methods. There are several factors determining the sensitivity. For example each method detects its own subset of interactions which is called the "assay sensitivity". Hence, sensitivity can be increased by combining different PPI detection methods (Sanderson, 2009; Stelzl and Wanker, 2006; Suter et al., 2008). Notably, this also holds true for different versions of the Y2H system. Provided that different Y2H systems are producing high precision data, parallel use of several Y2H setups will increase sensitivity which increases the workload (Figeys, 2008; Rajagopala et al., 2009; Schwartz et al., 2009). The "sampling sensitivity" is the fraction of all identifiable interactions that are found in a single trial of an assay performed. The sampling sensitivity in a Y2H-matrix screen is 45 % and approximately six screens are needed to reach 90 % saturation resulting in six times the workload (Venkatesan et al., 2009). Here, we describe a new workload reduced Y2H approach for comprehensive PPI mapping utilizing second generation sequencing.

4.1.1 Development of a new Y2H second generation sequencing screen

The overall aim is to reduce the workload and increase the sensitivity of the Y2H screen whilst producing high-quality PPI data. In Figure 6A and B the new Y2H second generation sequencing (Y2H-seq) screen is schematically presented in comparison to the standard Y2H-matrix screen, for one of four replicas. Ninety-six baits containing the proteins of interest were exemplarily tested for interactions against a set of preys here referred to as prey array. The prey array consisted of individually subcloned and characterized prey clones that expressed a particular prey protein in each well of a plate. In the Y2H-seq screen approach one prey array was pooled and subsequently split into four parts. Each of the parts was mixed with one of four pooled bait culture replicas. The mix was repeatedly spotted onto non-selective medium to allow mating. Hence, several thousand individual matings were performed. To select for interactions the mated yeast was transferred onto selective medium. Only diploid yeast containing a positive interaction pair of proteins can grow on selective medium. Plasmid DNA was isolated from grown yeast spots, prey plasmid inserts amplified and identified by second generation sequencing. Taking into account the number of obtained reads for each prey a ranked list of primary prey hits was generated.

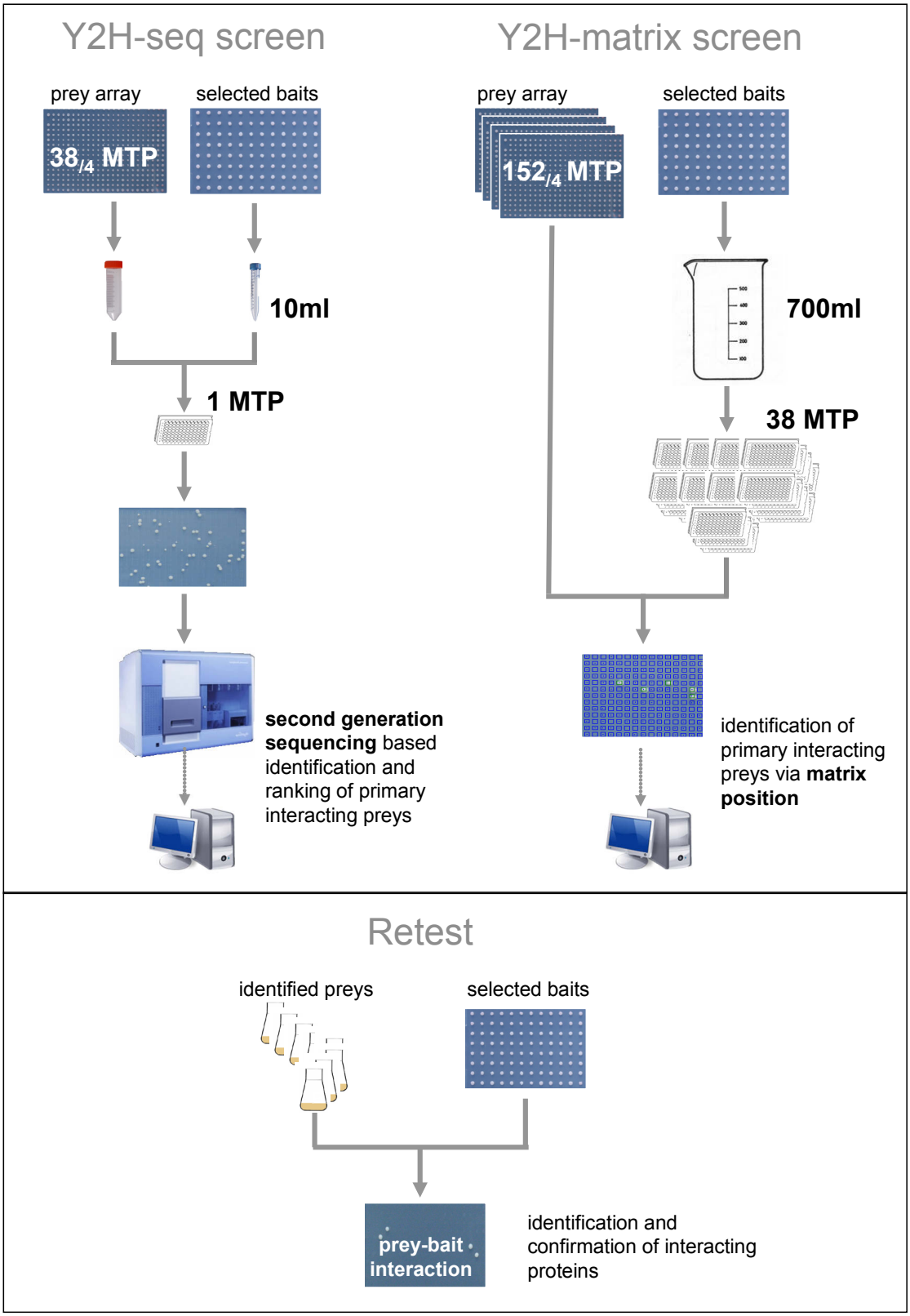

Figure 6: Schematic representation of the Y2H-seq and Y2H-matrix screen

The new Y2H-seq screen (left) and the standard Y2H-matrix screen (right) are schematically compared for one of four replicas. Selected baits transformed with the protein of interested were screened against the prey array containing 14,268 prey clones. Baits were separately grown and then pooled. In the Y2H-seq approach preys were separately grown and then pooled, too. Small volume of highly concentrated pooled preys was mixed with highly concentrated pooled baits. The mixture of preys and baits was distributed into one microtiterplate (MTP). The MTP was stamped 36 times on non-selective medium. Hence, 13,824 individual matings were performed. After mating the diploid

yeast was transferred to selective medium. Diploid yeast containing interacting bait and prey can grow on selective medium. Plasmid DNA of the yeast colonies was isolated, prey inserts amplified, fragmented and sequenced in a Solexa HighSeq2000 sequencer. On the right hand the Y2H-matrix approach. Here, a large volume of pooled bait culture were needed to fill 38 MTP and mate every well separately with one of the 14,268 preys in the ordered array. After mating the diploid yeast was transferred to selective medium. Interacting preys and baits grow on selective medium. In the Y2H-matrix screen preys interacting with at least one bait of the pool were identified by their array position. The retest at the bottom is identical for both approaches. The identified preys were grown in flasks and separately mated against each replica of the baits. Interacting preys and baits grew on the selective medium.

In the Y2H-matrix screen every prey was tested separately against the pool of baits. Hence, every prey was mated with the pool of baits in a specific position in the prey array and the interacting preys were identified by their array position. The mating was performed in four replicas and therefore four ordered prey arrays were needed. By computational analysis a retest list containing primary prey hits was generated. In comparison to the Y2H-matrix workflow in the Y2H-seq approach the pool of preys was mated with the pool of baits, subsequently reducing the workload. Hence, the workload in the Y2H-seq approach is reduced.

The primary prey hits identified in the Y2H-matrix and in the Y2H-seq screen were verified in a retest approach (Figure 6C). In the retest preys were tested separately against selected baits. Thus, the interacting proteins were unambiguously identified. The retest layout was identical for both screens and hence resulted in the same high confidence PPIs.

In summary, the novel PPI Y2H-seq mating approach reduced the workload substantially in comparison to the standard Y2H-matrix approach whilst produces the same high confidence PPI data.

4.2 Y2H screens for protein methyltransferase (PMT) and protein demethylase (PDeM) interactions

A significant percentage of the human proteome encodes for protein methyltransferases (PMTs) that catalyze the transfer of methyl groups to substrates (Petrossian and Clarke, 2011). So far, only about hundred proteins are known to be methylated (Bedford and Richard, 2005; Boisvert et al., 2005a; Pahlich et al., 2006) (Swissprot). The overall number of PMTs suggests that more, if not many more, proteins are methylated. However, identification of new substrates of methyltransferases is difficult because most methods require prior knowledge of the substrate identity (Ong et al., 2004; Rathert et al., 2008) or have other limitations such as lack of affinity reagents or high-quality antibodies (Komyod et al., 2005; Pahlich et al., 2006). Since the nature of methyltransferase activity depends on direct physical encounters between enzyme and target protein, methylation substrates can be identified by studying PMT-protein interactions (Passos et al., 2006a). Demethylases (PDeM) remove methyl groups and therefore interact with methylated proteins. Hence, we also studied PDeM-protein interactions to identify methylation substrates. Those enzyme-substrate interactions often appear to be transient. Therefore, the Y2H method was

used as it is particular suitable for detecting transient interaction (Vinayagam et al., 2010; Yu et al., 2008).

4.2.1 Selection of PMTs and PDeMs to screen for interacting proteins

PMTs and PDeMs were cloned and shuttled into bait expression vectors to screen for interacting prey proteins. Enzymes selected and finally screened are illustrated in Figure 7. Different bait clones were used to increase the probability to detect PPIs and improve the quality of the interaction data by verifying the interactions by multiply clones (appendix Table 6).

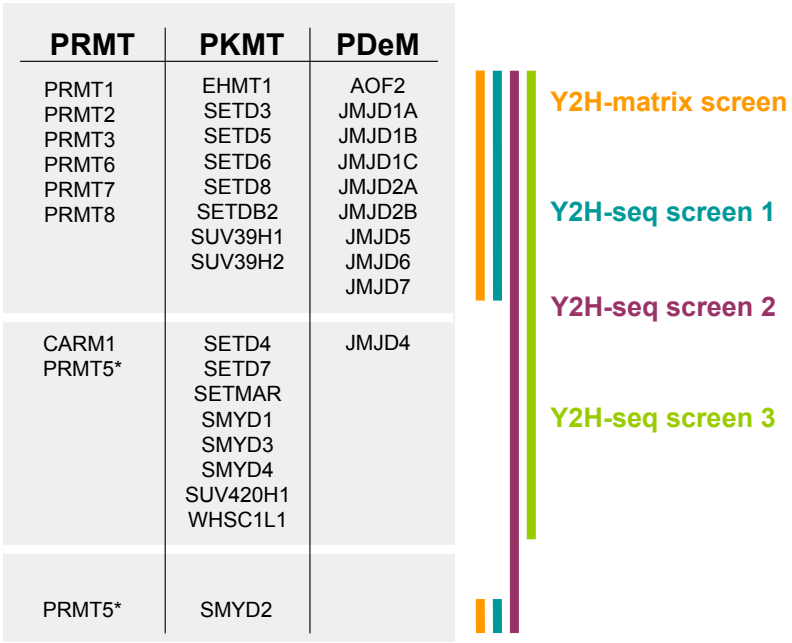

Figure 7: PRMTs, PKMTs and PDeMs used in the different screening approaches as baits

Enzymes are sorted by their enzymatic function: PRMTs (left) catalyze arginine methylation, PKMTs (middle) catalyze lysine methylation and PDeMs (right) remove methylation. The bars indicate baits used in the Y2H-matrix screen (orange), Y2H-seq screen 1 (blue), Y2H-seq screen 2 (purple) and Y2H-seq screen 3 (green). The asterisk marks the domain clone of PRMT5 screened in the Y2H-seq screen 2 and Y2H-seq screen 3, and the full length clone screened in the Y2H-matrix screen and Y2H-seq screen 1.

PRMTs are known to methylate non-histone proteins. Therefore, a comprehensive set of PRMTs was analyzed. This set includes PRMT1, the best studied PRMT, which methylates the majority of known non-histone substrates (Bedford and Clarke, 2009). Three PRMT1 clones were selected for an interaction screen. Also three PRMT2 clones were selected. PRMT3 was not available as gateway compatible entry clone in the laboratory collection. Hence, gateway compatible PRMT3 clones were generated in a 2-step gateway PCR reaction using specific primers followed by ligation into an entry vector. However, the generated PRMT3 clones lack the N-terminal region (amino acid 1 to 148) as primer hybridization was not possible due to adenine thymine stretches at the N-terminus. CARM1, also called PRMT4, was obtained as entry clone from the Incorporated Administrative Agency National Institute of Technology and Evaluation

(NITE). For PRMT5 two full lengths clones and one clone encoding for the amino acids 376 to 733 were chosen. One PRMT6 clone was available as entry clone in the laboratory collection and another one was cloned from a cDNA template. One clone was available for PRMT7 and three for PRMT8. These PRMT8 clones lack the N-terminal domain (amino acid 1 to 60).

SET domain containing methyltransferases catalyze methylation of lysine residues mainly on histones and ribosomal proteins. Only a few other substrates are known (Dillon et al., 2005; Webb et al., 2008). We chose 21 SET domain containing PKMTs out of 56. This includes EHMT1 (GLP) and EHMT2 (G9a) homologous histone methyltransferases whereof EHMT2 is known to methylate non-histone proteins (Rathert et al., 2008). In addition, SETDB1, SETD8, SETD7 and SETMAR were chosen as their methylation activity was previously investigated. Also putative protein lysine methyltransferases were chosen including SETD3, SETD4, SETD5, SETD6 and SETDB2 (Fritsch et al., 2010; Kurash et al., 2008; Lee et al., 2005d; Li and Kelly, 2011; Petrossian and Clarke, 2011). SETD7 was not available as entry clone. Thus, we generated full length and domain entry clones (Nishioka et al., 2002; Wang et al., 2001). Furthermore, we selected MYD and SET domain containing SMYD1, SMYD2, SYMD3 and SMYD4. SUV39H1, a well known histone modifier, and the closely related SUV39H2 were selected (Fritsch et al., 2010; O'Carroll et al., 2000; Rea et al., 2000). For WHSC1L1 methylation activity was demonstrated and therefore included (Kim et al., 2007).

In order to maintain a more comprehensive set of proteins involved in methylation we also selected demethylases (PDeMs). The best studied PDeM AOF2 and nine additional PDeMs containing the JmjC domain were chosen. These include JMJD1A, JMJD1B, JMJD2A and JMJD2B encoding PDeMs known to act on histones and JMJD1C, JMJD4, JMJD5 and JMJD7 which are putative PDeMs (Agger et al., 2008). JMJD6 was defined as histone arginine PDeM (Chang et al., 2007) and thus selected. Later it was characterized as lysyl-hydroxylase (Webby et al., 2009).

Selected clones were shuttled into a bait vector with an N-terminal lexA DNA-binding domain (pBTM116). Clones without a stop codon were additionally shuttled into a second bait vector with a C-terminal lexA DNA-binding domain (pBTMCc24) (Table 2). The shuttle reactions were transformed into *E. coli* expression strains. The clones were enzymatically digested to control for correct insert size. Additionally, the enzyme clones were Sanger sequenced.

In summary, a comprehensive set of PRMTs and large sets of PKMTs and PDeMs were successfully selected and shuttled into the appropriate bait vectors.

4.2.2 Preparation of PMTs and PDeMs for a Y2H PPI search

In order to perform a Y2H screen MATa yeast strains were separately transformed with bait plasmids using a high efficient yeast transformation protocol. Four replicas were generated. To detect autoactivation the bait strains were mated with a control prey strain carrying a prey plasmid without insert. Prior to screening it was important to remove autoactive bait strains because they

grow independent of interactions on selective medium after mating and mask any interaction signals of other baits in the pool. Yeasts colonies growing on selective medium after mating with the control prey strain were determined to be autoactive and removed. The clones of SETDB1, EHMT2, SETD7 full length (FL) and JMJD2A were identified as autoactive and removed. Additionally, after the second Y2H-seq screen PRMT5 FL clone and SMYD2 clone were removed due to their promiscuous interaction pattern. Hence, only SETD7 and PRMT5 domain clones were used in the screens instead of full length clones. In Table 1 autoactive and promiscuous baits are listed. We generated 82 MATa bait strains comprising 34 different enzymes in four replicas.

Table 1: Autoactive and promiscuous baits

PRMT	PKMT	PDeM
PRMT5 FL*	EHMT2	JMJD2A
	SETD7 FL*	
	SETDB1	
	SMYD2	

* Full length clones were autoactive.

Autoactive preys were only removed in the Y2H-seq screen approach because autoactive preys grow interaction-independent on selective medium and mask any interaction signals of other preys in a pool. Autoactive prey strains are repeatedly discovered in several screens performed in the laboratory with different sets of baits.

4.2.3 Overview of PMT- and PDeM-protein interaction screens

The 84 yeast strains expressing the PMTs and PDeMs were screened against a prey array expressing 12,199 different preys. This prey array was used in the Y2H-matrix as well as in the Y2H-seq screen and shared with the Wanker Laboratory (Max-Delbrück-Centrum for Molecular Medicine). The collection of 14,268 prey clones (representing 12,199 proteins) included most of the currently available human full-length cDNAs and covered over 57 % of the human protein coding genes.

Our aim was to identify PMT- and PDeM-interacting proteins. For this purpose non-autoactive PMTs and PDeMs were used in proteome-wide Y2H screens. One Y2H-matrix screen and three Y2H-seq screens were performed with slightly different sets of PMTs and PDeMs for technical reasons (Figure 7 and appendix Table 6). We screened 72 bait strains by the standard Y2H-matrix approach (Table 2). This set of baits was also screened in the Y2H-seq screen 1 for direct comparison. It contained 23 enzymes associated to methylation and in addition a set of unrelated baits including human immunodeficiency virus genes, tuberculosis genes, the STUB1 gene and the CHERP gene. Those unrelated baits were not considered in the analysis. In the Y2H-seq screen 2 we tested an enlarged set of 35 enzymes. PRMT5 full length and SMYD2 bait strains were identified as promiscuous and removed in the subsequent Y2H-seq screen 3.

In summary, we screened a large set of bait strains containing 35 PMTs and PDeMs against a prey array containing more than 14,000 prey clones.

Table 2: Detailed breakdown of the number of baits used

Screen	Baits in pBTM116	Baits in pBTMCc24	Unrelated baits	Bait strains	PMTs and PDeMs
Matrix	39	0	33	72	23
Seq1	39	0	33	72	23
Seq2	47	37	0	84	35
Seq3	45	37	0	82	34

4.2.3.1 Proteome-wide screen of PMTs and PDeMs in a Y2H-matrix approach

To identify PMT- and PDeM-protein interactions bait strains expressing the enzymes were individually screened against 14,268 preys in the Y2H-matrix approach. More than half a million pairwise interaction tests are needed to screen each of the 39 baits separately against each prey in the array. Hence, the bait clones were pooled and therewith screening effort reduced. Thirty-nine baits were grown separately and pooled just before mating. This was repeated for each of the four replicas separately. The pool of each replica was pipetted into 38 microtiterplates (MTPs) and tested separately against every single prey of the 38 MTPs of the array. Yeast mating was performed on non-selective medium (YPD). To detect interactions between the pool of baits and preys diploid yeast was transferred onto selective medium. Preys interacting with the pool of baits grew on selective medium and plates were photographed. In Figure 8 selective growth of the pooled baits mated with 384 preys is shown. Of those 384 matings four yeast spots grew on the selective medium indicating a potential interaction (primary hit) between the prey and at least one bait of the pool.

Primary hits identified in the Y2H-matrix approach
Selective growth of preys interacting with at least one bait of the pool

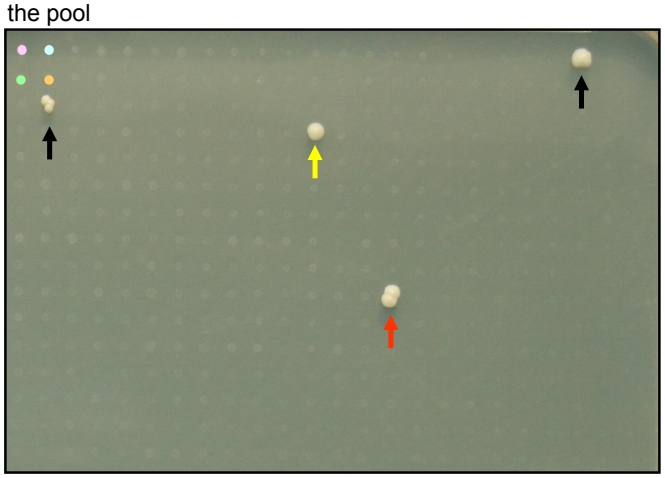

Figure 8: Selective growth of primary hits in the Y2H-matrix approach

Growths of diploid yeast spots on selective medium indicate interactions. Dots in the upper left corner indicate spot pattern of the MTP: A1 (pink), A2 (blue), B1 (green) and B2 (orange) each containing

an individual prey strain and the pool of bait strains. One replica of the MTP eight of the 38 array plates is exemplary shown. Growth of the yeast colony in position D12 (yellow arrow) indicates an interaction between the prey strain containing the PPARA protein and at least one bait of the pool. In position J15 (red arrow) the prey strain containing TMED5 interacts with at least one bait of the pool. A22 and C2 (black arrows) point towards autoactive prey strains.

The interacting preys were identified by their position in the prey array. The Visual Grid software was used to obtain the position of every grown spot and assign it to the individual prey strains in the array. In the Y2H-matrix screen 2158 yeast spots of different intensities were identified. Small yeast spots were assigned to an intensity of one and bigger colonies to an intensity of three. In principle, all bait pool/prey combination that result in a growth signal should be retested. However, there is a trade-off between the absolute number of interaction recorded and the retest success rate (Worseck et al., 2012). Therefore, bait pool/prey combinations that grown in at least two out of the four copies were considered to be retested (537) (Table 3). Of these prey strains 128 were autoactive and 217 had a background signal of two small colonies with an intensity of one. These were removed from the retest list. In the end, the Y2H-matrix screen when performed with 72 baits according to our standard protocol (Worseck et al., 2012) revealed 192 primary hits to be retested for PPIs (Table 3).

Table 3: Primary hits of the Y2H-matrix screen

Primary hits	No.
Total hits	2158
Hits come up > 2	537
Non autoactive hits	409
Hits intensity >2* for retest	192

A retest was performed to verify the interacting prey proteins and identify the bait proteins interacting with the prey.

4.2.3.2 Proteome-wide screen of PMTs and PDeMs applying Y2H-seq

Next we wanted to identify preys interacting with the PMTs and PDeMs using the novel Y2H-seq approach. In the first Y2H-seq screen the same 39 bait were screen as in the Y2H-matrix screen. Additionally, an enlarged set of 84 and 82 baits were screened in the second and third Y2H-seq screen, respectively (Figure 7 and appendix Table 6). Bait strains were grown separately on agar plates, scraped off and then pooled as in the Y2H-matrix screen. In the Y2H-seq screen the 14,268 preys were grown separately and then pooled, too. The prey pool was split into four parts and mated with each of the four replicas of the bait pools. The pool bait/pool prey combinations were repeatedly stamped (13,824) on non-selective medium. This allowed individual matings in over ten thousand position. To detect interactions the mated pool bait/pool prey spots were transferred to selective medium. To improve selection the diploid yeast spots were stamped in a high dense pattern onto selective agar (Figure 9). Yeast spots grown on selective medium indicated

interactions of at least one prey of the pool with at least one bait of the pool. Noticeably, many more spots grew on the selective agar in the Y2H-seq screen in comparison to the Y2H-matrix screen (compare Figure 8 and Figure 9) suggesting that more interactions were detected.

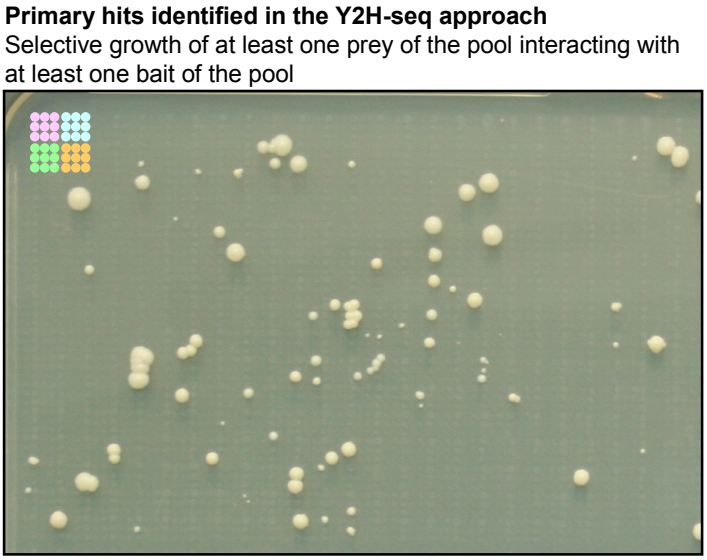

Figure 9: Selective growth of primary hits in the Y2H-seq approach

Growth of diploid yeast spots on selective medium indicating interactions. Dots in the upper left corner indicate spot pattern. A1 (pink), A2 (blue), B1 (green) and B2 (orange) were stamped each nine times on the agar, each containing the pool of preys and the pool of baits. Grown spots indicate interaction between at least one prey of the pool with one bait of the pool. Prey strains interacting were identified by second generation sequencing.

The interacting preys were not identified over their position because all preys were pooled and every spot potentially contains all possible preys. To identify the interacting preys, prey plasmid DNA was extracted from the yeast spots growing on the selective medium and sequenced. Therefore, all yeast spots were scratched of the agar. Plasmid DNA was isolated from the yeast. In Figure 10A a defined band for the plasmid DNA of the four replicas A, B, C and D is shown. The insert coding for the interacting preys was amplified by PCR with prey-vector specific primers. The primers aligned to the prey vector sequence N-terminal and C-terminal of the insert. Thus, prey inserts from a few hundred to a few thousand base pairs lengths were amplified and appear on an agarose gel as a smear of DNA (Figure 10B).

After further amplification and fragmentation the probes were sequenced using the Illumina Genome Analyzer (by the second generation sequencing team, Bernd Timmermann, MPIMG in Berlin). The obtained reads were mapped to NCBI Reference sequence (RefSeq) annotated genes. Identified proteins were ranked by the quantity of reads in the four replicas (in cooperation with Jonathan Woodsmith and Arndt Grossman, OWL Stelzl, MPIMG in Berlin). The list of ranked prey proteins contained more than 7,000 entrez gene IDs. However, the coverage for each clone was not uniform with on average 72 % of mapped genes in each run having fewer than five reads mapped to it. Furthermore, 99 % of the total reads form each screen mapped to the top 300 ranked

genes.

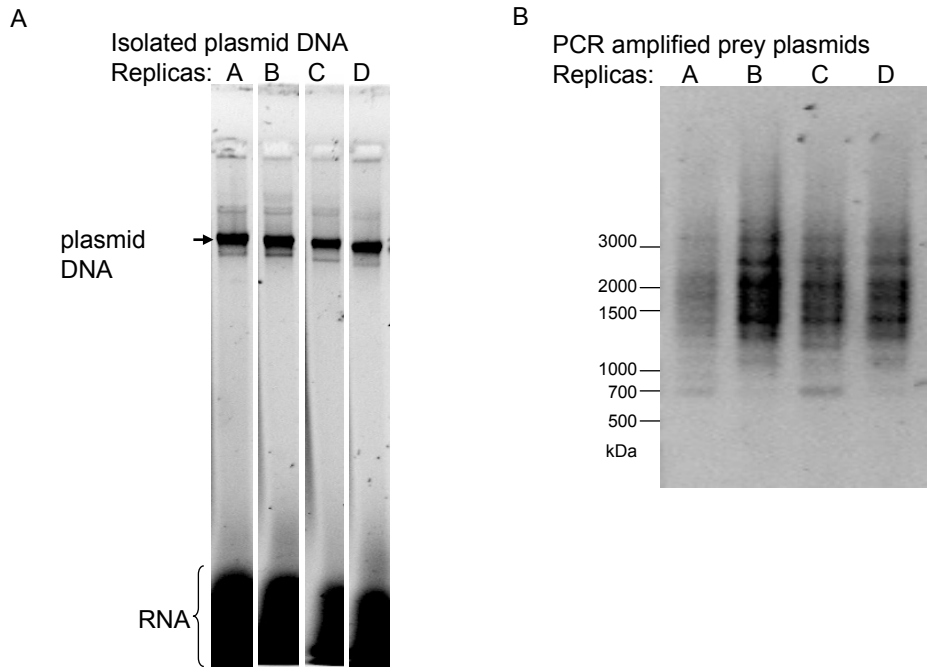

Figure 10: Prey plasmid preparation for second generation sequencing.

Preparation of prey plasmid DNA of the four replicas A, B, C and D. (A) Isolated plasmid DNA from colonies grown on selective medium on 1 % agarose gel. (B) The prey plasmids were amplified from the isolated plasmid DNA with specific primers. The amplified prey plasmid DNA of various lengths are seen as a smear on 1 % agarose gel. The amplified prey plasmid encodes for the proteins interacting with the pool of baits.

In summary, we reduced the screening effort by pooling the preys and resulted in many more spots grown on selective medium. Those were collected and primary hits identified by sequencing the prey plasmid of the yeast spots grown on selective medium. The retest list was generated by ranking the preys according to their quantity of reads. To verify primary hits and identify the interacting bait protein a retest was performed.

4.2.4 Verification of primary hits identified in the Y2H-matrix and Y2H-seq screens

In the Y2H-matrix screen 192 preys fulfilled the rested criteria and were selected to be retested. The top 240 identified preys of the first Y2H-seq screen were retested. Of the second and third Y2H-seq screen the top 284 and top 401 preys were retested, respectively. As the retest was identical for both screening approaches preys identified in more than one screen were in general only retested once. Retest was only repeated for specific control experiments. Altogether more than 800 preys were retested separately against the bait proteins.

The identified preys were tested against the four replicas of the baits to identify the bait protein interacting with the prey. Preys were grown in flasks and mated with the bait proteins on non-selective medium. The individually mated bait and prey were stamped on selective medium and interaction detected by growth. This way each yeast spot corresponded to one bait/prey

combination only. In Figure 11 examples of yeast spots verifying and deconvoluting the interactions between bait-prey pairs are shown for selected interactions involving different PMTs. PRMT1 interacted with splA/ryanodine receptor domain and SOCS box containing 2 protein (SPSB2). PRMT2 interacted with Pre-mRNA 3'-end-processing factor FIP1 (FIP1L1). CARM1 and PRMT6 interacted with Pygopus homolog 1 (PYGO1). PYGO1 has a nuclear localization signal and a PHD like zinc finger which binds histone H3 at methylated lysine 4 (Fiedler et al., 2008).

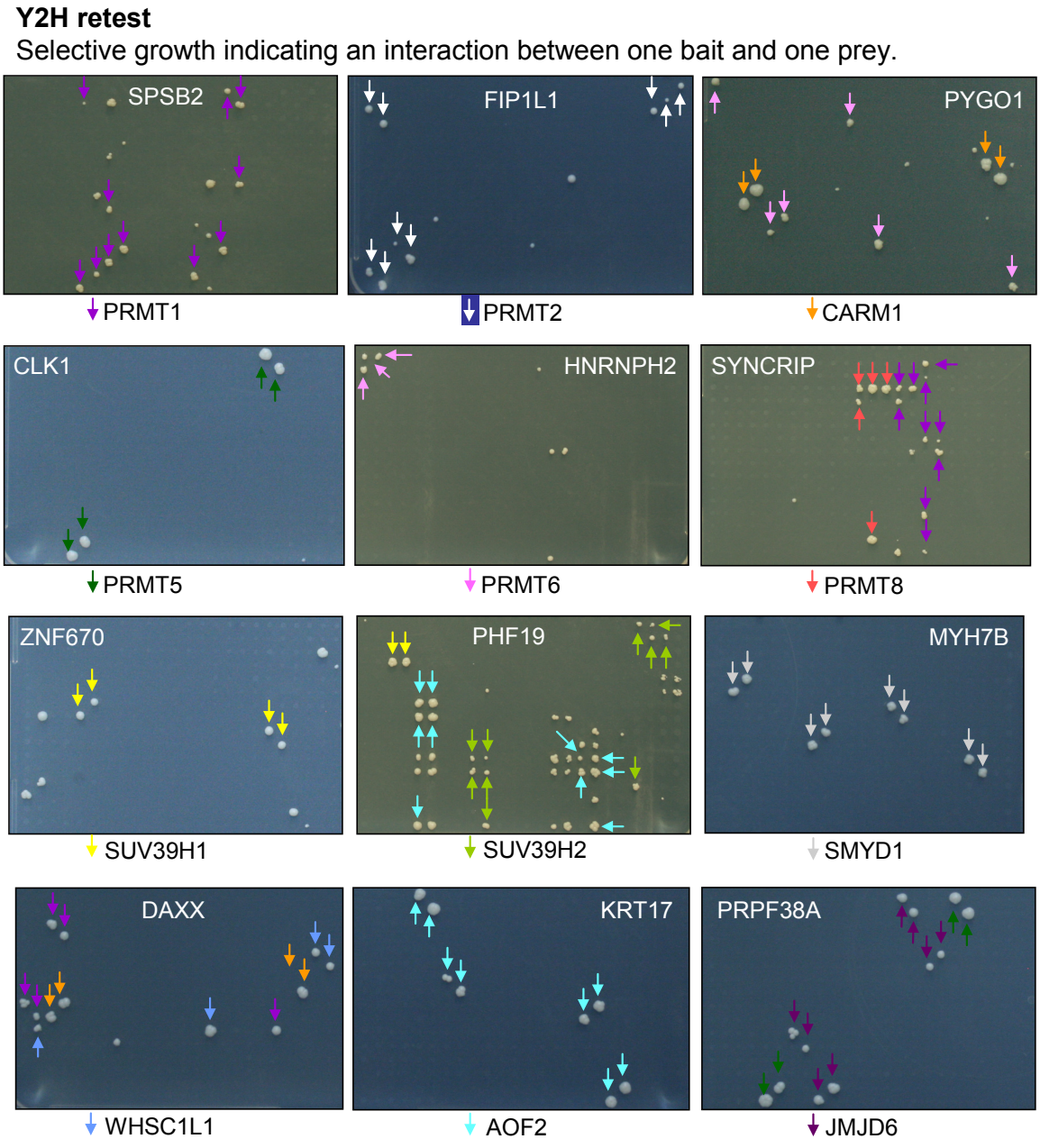

Figure 11: In the retest verified interactions between bait and prey proteins

Selective Y2H plates of a retest experiment are exemplarily shown. Different prey proteins (white) were tested against a 384 MTP containing the bait proteins. Growing yeast colonies verify interaction between bait-prey pairs. Baits are indicated by colored arrows.

PRMT5 bound to the family of dual specificity serine/threonine kinases CLK1 and CLK3 (not

shown). Those kinases regulate the intranuclear distribution of the serine/arginine rich (SR) family of splicing factors (Duncan et al., 1998). PRMT6 interacted with heterogeneous nuclear ribonucleoprotein (hnRNP) HNRNPH2 which is predicted to be methylated in the Swissprot database. SYNCRIP also a hnRNP protein known to be methylated (Araya et al., 2005; Kim et al., 2008; Ong et al., 2004; Passos et al., 2006b) interacted with PRMT1 and PRMT8.

SUV39H1 interacted with a number of zinc finger containing protein including ZNF670. SUV39H2 interacted with PHD finger protein 19 (PHD19). PHD fingers are known to bind methylated proteins (Corsini and Sattler, 2007). The heavy chain of myosin II encoded by MYH7B interacted with SMYD1. WHSC1L1 interacted with death-domain associated protein (DAXX). DAXX interacted also with CARM1 and PRMT1. AOF2 interacted with many keratins including KRT17. JMJD6 interacted with the splicing factor PRP38 pre-mRNA processing factor 38 (yeast) domain containing A (PRPF38A). PRPF38A is important for stabilization of U6 small nuclear RNA levels (Blanton et al., 1992).

All together we verified and deconvolute 523 interactions of 324 prey proteins interacting with 22 bait proteins (Figure 23).

4.3 Comparison of the Y2H-matrix and the Y2H-seq approaches

4.3.1 Comparison of the preys identified in the Y2H-matrix and the Y2H-seq screens

We screened the same set of bait proteins in the Y2H-matrix and in the Y2H-seq screen 1. Hence, we expect a significant overlap of interacting preys proving that the new Y2H-seq approach produces interaction data comparable to the Y2H-matrix approach.

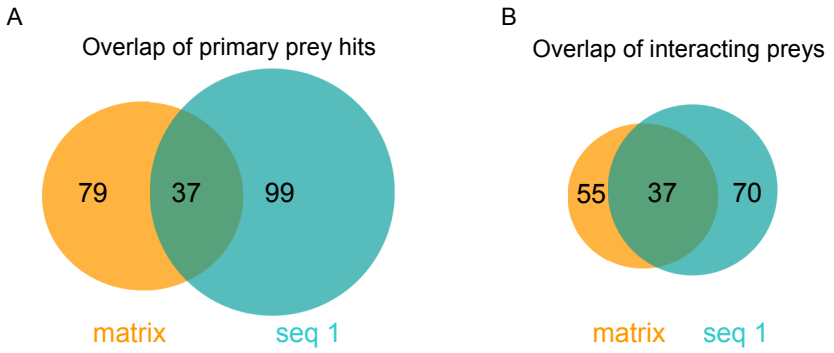

Figure 12: Overlap between Y2H-matrix and Y2H-seq screen 1

(A) The overlap between the 116 primary hits identified in the Y2H-matrix screen (orange) and the 136 primary hits identified in the Y2H-seq screen 1 (blue) is illustrated in a Venn diagram. (B) The Venn diagram illustrates positively retested interacting preys overlapping between the Y2H-matrix screen (orange) and the Y2H-seq screen 1 (blue).

Of the 192 primary hits identified in the Y2H-matrix screen 76 interacted with unrelated baits. In the first Y2H-seq screen 104 of the top 240 primary hits identified interacted with unrelated baits. We compared the 116 primary Y2H-matrix hits with the 136 primary hits of the

Y2H-seq screen 1 interacting with PMTs and PDeMs. Thirty-seven primary hits overlap between the Y2H-matrix and the Y2H-seq screen (Figure 12A). Similar overlap is expected if the Y2H-matrix screen is repeated (Worseck et al., 2012). Hence, the primary hits identified in the Y2H-seq screen were comparable to the primary hits identified in the Y2H-matrix approach.

In the Y2H-matrix screen 92 primary hits and in the Y2H-seq screen 107 primary hits were positively retested. Thirty-seven of those overlap between the Y2H-matrix and the Y2H-seq screen (Figure 12B). Hence, primary preys identified in both screens were true interacting preys.

4.3.2 Comparison of the bait interactions in the Y2H-matrix and the Y2H-seq screens

To compare the number of interactions per bait between the screens an interaction rate for each bait was calculated. The interaction rate was calculated as percentage of total interactions found in the screen (Figure 13).

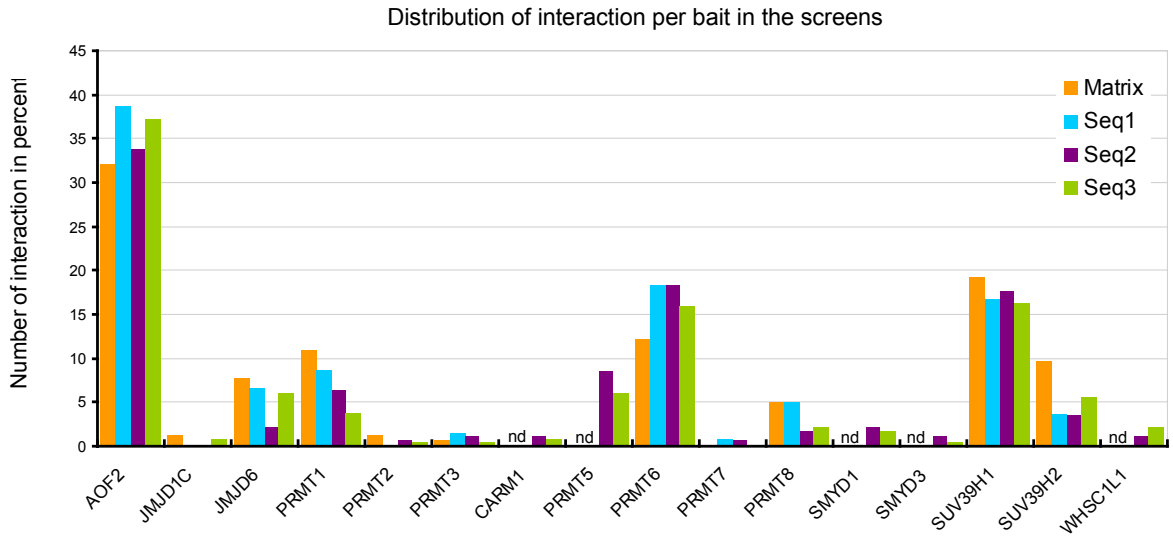

Figure 13: Comparison of interaction per bait in different screening approaches

Interaction rate is indicated on the y-axis corresponding to the different baits on the x-axis. Colors stand for different screens (see legend). Interaction rate was calculated from the 192 primary hits of the Y2H-matrix screen and the top 192 ranking preys of the Y2H-seq screens. CARM1, PRMT5, SMYD1, SMYD2 and WHSC1L1 were not tested in the Y2H-matrix and Y2H-seq screen 1 and therefore not determined (nd).

In all screens around 32 % of the preys interacted with AOF2. Hence, AOF2 had in all screens a high interaction rate like wise SUV39H1. In contrast, PRMT2 had a low interaction rate in all screening approaches. Also the interaction rates of JMJD1C and PRMT7 were constantly low in all screens. Hence, if the interaction rate was low in one screen it was also low in the other screens. PRMT6 interaction rate was about 10 % lower in the Y2H-matrix screen compared to the Y2H-seq screens. The opposite effect was visible for JMJD6 and SUV39H2. The variations observed were small and they seem not to be dependent on the Y2H-matrix or Y2H-seq approach in general.

In summary, the bait interactions were constant in the different screening approaches. Those

results showed that the Y2H-seq approach detected interactions for baits interacting with a few preys equally well as baits interacting with many preys.

4.3.3 Evaluation of the retest success rates in the Y2H-matrix and the Y2H-seq screens

The retest success rate is the percent of primary prey hits which interact with at least one bait. To compare the retest success rate between the different screens the 192 primary hits of the Y2H-matrix screen and the top 192 preys of the Y2H-seq screens were considered. Of the 192 preys identified in the Y2H-matrix screen 92 interacted with PMTs or PDeMs whereas 24 preys did not interact at all. The remaining 76 preys interacted with unrelated set of baits. Thus, the retest success rate was 0.79. Of the top 192 preys identified in first Y2H-seq screen 92 preys interacted with PMTs or PDeMs but only 13 preys did not interact and this resulted in a retest rate of 0.88. In the second and third Y2H-seq screen the retest success rates accounted 0.85 and 0.83 for the top 192 primary hits, respectively (Figure 14).

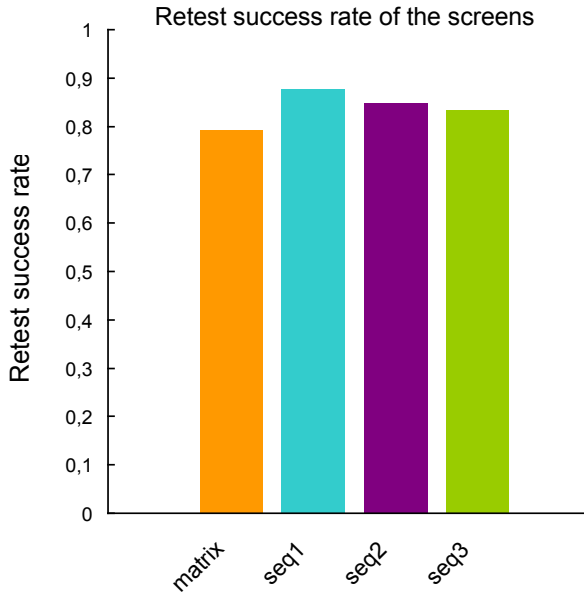

Figure 14: Retest success rates of the screens performed

Retest success rate of the 192 primary hits in Y2H-matrix screen and of the top 192 primary hits in the Y2H-seq screens. Retest success rates were calculated by dividing PMT interacting preys by non interacting preys. Preys interacting with non related baits were not considered.

If identical criteria were applied we reached similar a high retest success rate above 79 % in the screens. This is only the case if all 192 hits obtained in the Y2H-matrix approach were compared to the top 192 hits obtained in the Y2H-seq approaches.

4.3.4 Strong correlation of retest success rate and Y2H-seq prey rank

The primary prey hits of the Y2H-seq screens were ranked using a score generated from the number of sequences mapped to each prey sequence. Hence, the top ranks were identified by many

reads. This let to the assumption that the top ranked genes have a higher retest success rate, i.e. were more likely to be true interacting proteins, than genes identified with smaller numbers of reads.

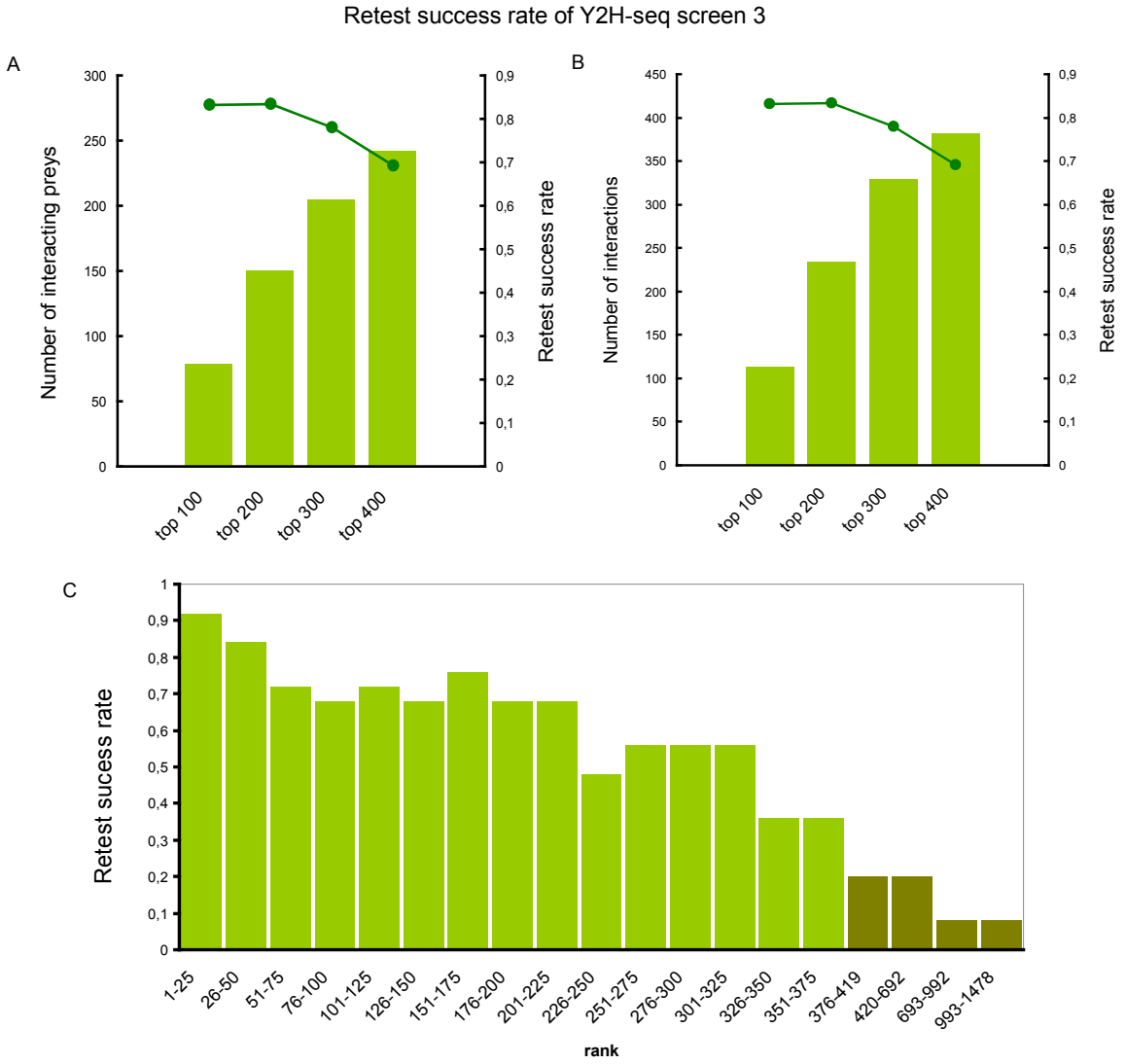

Figure 15: Retest success rate of Y2H-seq screen 3

Top 400 retested preys of the Y2H-seq screen 3. (A) Left axis indicates the number of confirmed interacting preys. Right axis indicates the success rate of retested preys. (B) Left axis indicates the number of confirmed interaction. Right axis indicates the success rate of retested preys. (C) Retest success rate was calculated in bins of 25 preys from rank 1 to 375 (light green). Ranks above 375 were not tested continuous. Thus, the retest success rate above rank 375 was calculated from the next bin of 25 preys tested (dark green). Unrelated or promiscuous bait interactions and autoactive preys were excluded from the calculation of the retest success rate.

The retest success rate of the top 100 identified preys in the third Y2H-seq screen was 0.83. The retest success rate of the top 400 preys retested drops to 0.69 (Figure 15A). A similar picture emerged if interactions instead of proteins were considered (Figure 15B). In the top 100 retested preys 113 interactions were detected as some of the identified preys interacted with more than one

bait. Altogether, 382 interactions were identified in the third Y2H-seq screen. A similar inverse correlation between the decrease of the retest success rate and ranks is also apparent in Figure 15C. The retest success rate was calculated in bins of 25 preys. The retest success rate of the top 25 ranks was 0.92. Up to the rank 325 the retest success rates were above 0.5 and then the retest success rate dropped to 0.3. In control experiments preys were retested at random. The 25 preys retested between rank 401 and 618 had a success rate of 0.25. Then the retest success rate of the randomly tested preys drops to 0.1. Hence, in the Y2H-seq approach top ranking preys obtained by many reads are retested with a higher retest success rate resulting in a quantitative relation ship between the number of obtained reads and the likelihood of true interactions.

4.3.5 Increased sensitivity through Y2H-seq screening

In the previous section it was shown that the primary hits identified in the Y2H-seq screen were comparable to the primary hits identified in the Y2H-matrix approach (4.3.1). Also interaction rates of the baits between both approaches were comparable (4.3.2). Furthermore, in the Y2H-seq screen the primary hits could be ranked to there quantity of reads (4.3.4). Hence, we could calculate retest success rates of the top 100, top 200 and so forth. Therefore, we were able to compare at the same retest success rate the number of interacting preys between the Y2H-matrix and Y2H-seq approach.

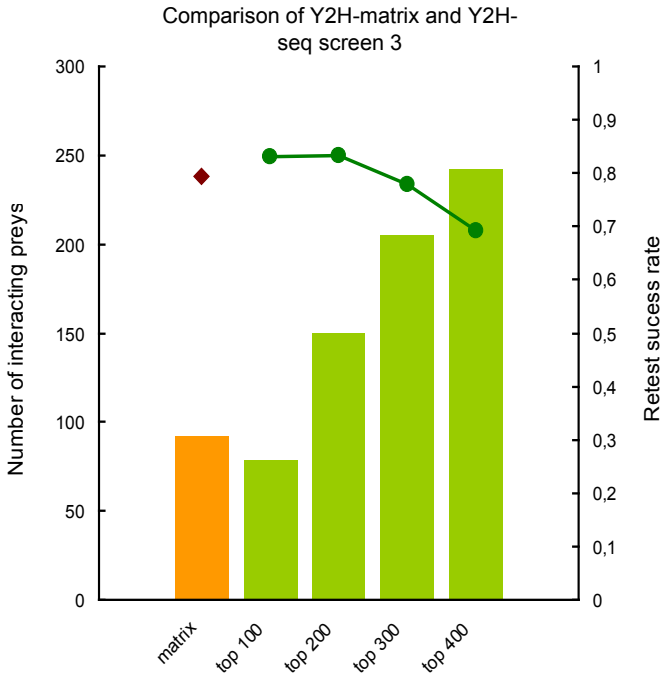

Figure 16: Comparison of the retest success rates of Y2H-matrix and Y2H-seq screen 3

Bars indicate number of interacting preys in Y2H-matrix (orange) and in the Y2H-seq screen 3 (green). The diamond indicate the retest success rate of the 192 preys in the Y2H-matrix screen whereas the circles in green show the retest success rates of top 100, top 200, top 300 and top 400 preys identified in the Y2H-seq screen 3. Unrelated or promiscuous bait interactions and autoactive preys were excluded from the calculation of the retest success rate.

The retest success rate of the 192 primary hits in the Y2H-matrix approach was 0.79 which was comparable to the retest success rate of 0.78 of the top 300 primary hits in the Y2H-seq approach (Figure 16). Hence, in the Y2H-seq approach more primary hits were tested at a similar retest success rate and thereby more PPIs were identified. Many more matings were performed in the Y2H-seq screen compared to the Y2H-matrix approach (4.2.3.2). This resulted in many more spots grown on the selective medium (compare Figure 8 and Figure 9). This and that more primary hits were tested at a similar retest success rate implied that the sensitivity of the Y2H-seq is increased. As we use the same version of the Y2H system we conclude that not the assay sensitivity is increased but rather the sampling sensitivity.

All further analysis were performed on the complete set of interaction detected in the Y2H-matrix, Y2H-seq screen 1, Y2H-seq screen 2 and Y2H-seq screen 3.

4.4 Interactions of PMT and PDeM proteins

4.4.1 Distribution of interactions across PMTs and PDeMs

While the focus was on methodological aspects in the previous sections here the number of interactions per PMT and PDeM is discussed. In Figure 17 the number of autoactive, interacting and not interacting PMTs and PDeMs is indicated.

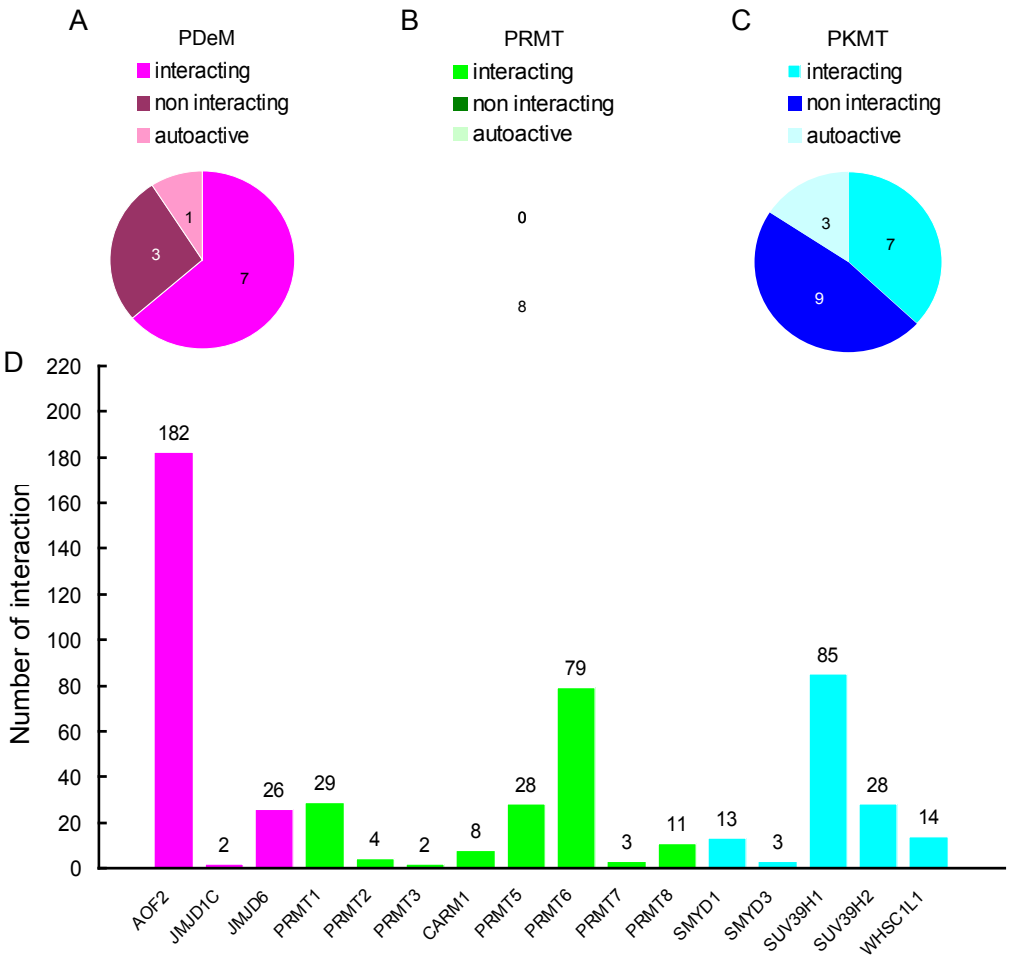

Figure 17: Number of interactions per PMT and PDeM

Demethylases indicated in pink, PRMT in green and PKMTs in blue. (A, B, C) The color nuance of the circles indicates interacting, not interacting and autoactive PMTs and PDeMs. (D) The diagram bars show the absolute number of prey proteins interacting with PMTs and PDeMs.

Of the demethylases a single enzyme was autoactive, three did not interact at all and seven interacted (Figure 17A). All PRMTs interacted with prey proteins (Figure 17B). Three PKMTs were autoactive or promiscuous, nine did not interact and seven interacted with prey proteins (Figure 17C).

16 PMTs and PDeMs interacted with more than one prey protein and are illustrated in Figure 17D. AOF2 is a well studied PDeM and was identified to interact with 182 prey proteins. We identified 26 prey proteins interacting with JMJD6. For the remaining JmjC domain containing PDeMs no interactions were identified except JMJD1C interacted with two prey proteins. For the minor studied PRMTs including PRMT2, PRMT3 and PRMT7 only few interacting prey proteins were detected. However, a larger set of interacting prey proteins was identified for the major studied PRMTs including PRMT1, CARM1, PRMT5, PRMT6 and PRMT8. Seven of the 19 PKMTs interacted with prey proteins. MYD domain containing PKMTs SMYD1 and SMYD3 interacted with prey proteins. SUV39H1 interacted with 86 prey proteins SUV39H2 in contrast only with 28. The active lysine methyltransferase WHSC1L1 interacted with 14 prey proteins.

As expected many PMTs and PDeMs interacted only with a few prey proteins and only a few PMTs and PDeMs interacted with many prey proteins (Barabasi and Oltvai, 2004). A representative interaction dataset was generated for AOF2 and JMJD6 as potential PDeMs, as well as for the major PRMTs and for the PKMTs including SMYD1, SUV39H1, SUV39H2 and WHSC1L1.

4.4.2 High interaction specificity of PMT and PDeM bait proteins

We screened a diverse set of enzymes involved in protein methylation including enzymes methylating lysine or arginine residues and enzymes removing methyl groups. Therefore, we expected that the baits have high substrate specificity and hence different enzymes interact with different proteins.

To analyze the substrate specificity of PMTs and PDeMs the overlap of interacting prey proteins between the different PMTs and PDeMs was examined. The heatmap in Figure 18 shows the percentage of prey proteins overlapping between the PMTs and PDeMs. If prey proteins interacted with one specific PMT or PDeM the overlap was zero (yellow square). If all prey proteins interacted with other PMTs or PDeMs the overlap was one (red square). For example AOF2 interacted with prey proteins which also interacted with SUV39H1 (orange square) but did not interacted with prey proteins interacting with PRMT8 (yellow square). The rate of unique prey proteins (i.e. only found with the respective PMT or PDeM) for every enzyme is shown in the top row. Small values (yellow square) pointed to a few prey proteins interacting uniquely with the bait

and high values (red square) to many prey proteins interacting uniquely with the bait. For example SMYD1 interacting prey proteins in general did not interact with other baits (top row yellow squares). Thus, it had a high value for the uniqueness of the prey proteins in the top row (top row red square). Noticeable, PRMT8 shared almost all of the interacting prey proteins with PRMT1 (red square) whereas PRMT1 shared fewer interacting prey proteins with PRMT8 (orange square). Hence, PRMT8 had a low number of prey proteins exclusively interacting with PRMT8. Thus, the value of unique prey proteins for PRMT8 was low (top row yellow square). PRMT1 in comparison had more unique prey proteins (top row orange square). SUV39H2 interacting prey proteins overlapped with SUV39H1 interacting prey proteins (orange square) whereas the other way around the rate of overlap was smaller (yellow square). None of the three prey proteins interacting with PRMT7 were unique resulting in high values in the PRMT7 row (red squares) and a zero for unique prey proteins in the top row (yellow square).

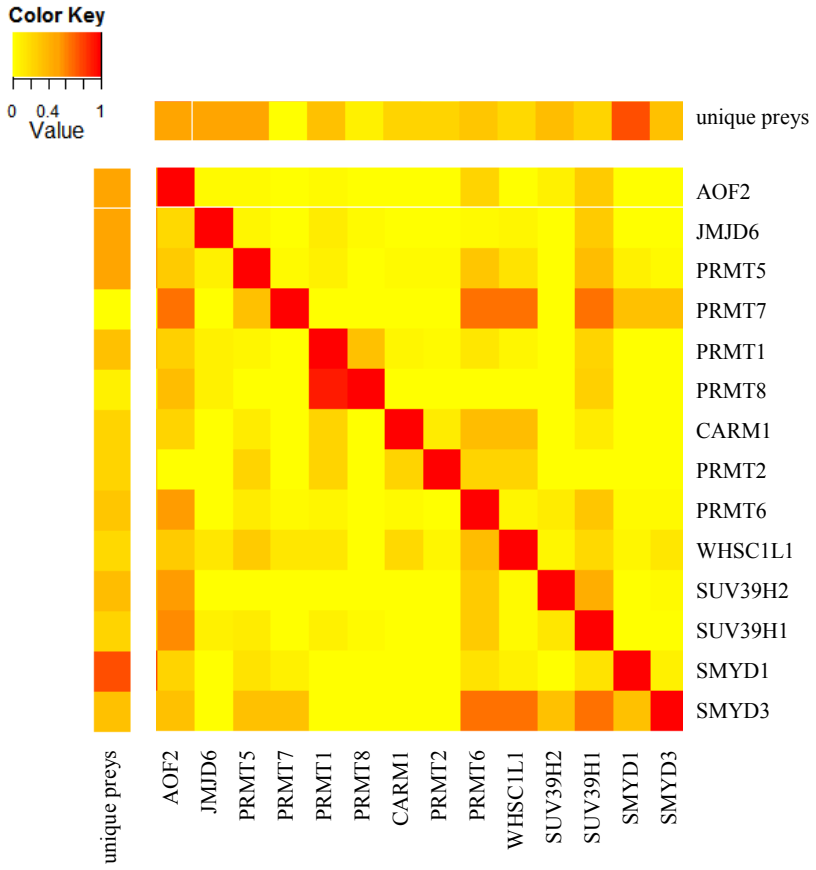

Figure 18: Heatmap of overlapping prey protein interactions between baits

The enzymes were ordered based on primary protein sequence alignment. The color indicates the rate of overlapping prey proteins between two enzymes proteins. This was calculated by dividing the number of preys interacting with both enzymes by the total number of preys interactions with the enzymes indicated in the column. Red indicates high and yellow low overlap of prey proteins interacting with both enzymes. The top row and the left column represent the uniqueness of the prey proteins. This indicates the number of prey proteins which were exclusively interacting with the bait. Red indicates high and yellow low rate of prey proteins exclusively interacting with the bait.

Most values in the heatmap were low (yellow and orange squares) indicating that interacting

prey proteins were not commonly shared between the baits. The top row had mainly high values (colored in orange and red) indicating that the prey proteins tended to interact with single PMT or PDeM. This pointed towards high specific PMTs and PDeMs interactions obtained in the Y2H system.

4.5 Analysis of PMT and PDeM interacting prey proteins

4.5.1 Known methylated proteins in the PMT and PDeM interaction dataset

Proteins interacting with modifying enzymes are likely to be substrates of those enzymes. Therefore, we first analyzed whether the proteins identified were posttranslationally modified. The uniprot database containing a total of 69 arginine methylated non-histone proteins and 25 lysine methylated non-histone proteins. Hence, the set of known methylated proteins was small. The 324 identified prey proteins were searched through for known methylated proteins and 11 proteins were indeed known to be methylated. To observe if methylated proteins were enriched in the dataset posttranslational modifications of the proteins identified in the screen were compared to posttranslational modifications of all proteins contained in the prey array. Figure 19 shows fold enrichment of PTMs over the prey array background. Proteins methylated on arginine residues were 2.7 fold enriched in the interaction dataset. Proteins methylated on lysines were 1.8 fold enriched but only two proteins methylated on a lysine residue were identified. Therefore, methyllysine enrichment was not significant. In comparison phosphorylated and acetylated proteins were not enriched.

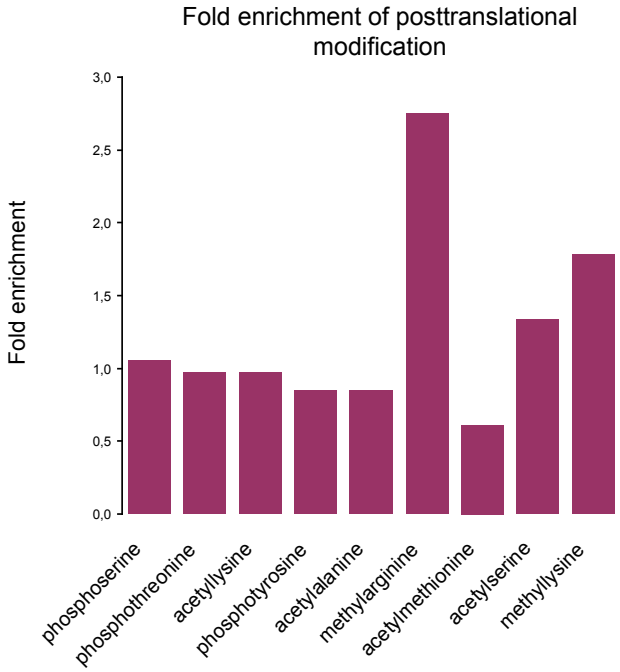

Figure 19: Posttranslational modifications on proteins identified in the Y2H screens

Fold enrichment of phosphorylation, acetylation and methylation in the PPI dataset over the prey array.

Six arginine methylated proteins were identified using the uniprot database and three by literature searches. In Table 4 known methylated proteins identified in the screens are listed. KRT7 encodes the type I intermediate filament chain keratin 7 and is methylated on arginine 19 (Ong et al., 2004). FUS encodes a multifunctional protein component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complex and is methylated on arginine 215 and 217 by PRMT1. Defects in FUS are the cause of amyotrophic lateral sclerosis type 6 (ALS6) (Kwiatkowski et al., 2009; Van Langenhove et al., 2010; Vance et al., 2009). Arginine methylation by PRMT1 participates in the nuclear-cytoplasmic shuttling of FUS, particularly of ALS6-associated mutants, and thus contributes to the toxic gain of function conferred by these disease causing mutations (Tradewell et al., 2012). We identified FUS as SUV39H1 interaction partner and KRT17 as AOF2 interaction partner. This does not corroborate the literature defined arginine methylation in both cases as both enzymes are characterized as lysine linked enzymes. FUS, HNRNPH2 and SYNCRIP belong to the large family of hnRNP proteins which are involved in pre-RNA processing and nucleocytoplasmic RNA transport. Several hnRNP proteins contain RG-rich regions which are known to be asymmetric dimethylated (Bedford and Richard, 2005; Liu and Dreyfuss, 1995). HNRNPH2 is predicted to be arginine methylated by swissprot. In agreement with the prediction we identified HNRNPH2 interacting with PRMT6. SYNCRIP was identified interacting with PRMT1 and PRMT8. SYNCRIP is known to be methylated in the RG-rich regions by PRMT1 and is thereby translocated into the nucleus (Passos et al., 2006b). Furthermore, we identified three RNA-binding proteins including the Ewing sarcoma breakpoint region 1 (EWSR1) protein identified interacting with PRMT1 and known to be methylated on several arginines by PRMT1 and PRMT8 (Pahlich et al., 2008). It is an RNA-binding protein containing a transcriptional activation domain and a RNA-recognition motif. If EWSR1 is methylated by PRMT1 the transcriptional activation of EWSR1 is down regulated through nuclear exclusion of EWSR1 (Araya et al., 2005; Kim et al., 2008). SERBP1 is also a RNA-binding protein and a paralog of SERBP1L. Both proteins are known to be methylated by PRMT1. We identified SERBP1 as PRMT8 interaction partner. SERBP1 is predominantly localized in the nucleus and upon demethylation it is partially redistributed to the cytoplasm (Passos et al., 2006a). Sam68 RNA binding protein is one of the best studied arginine methylation substrates of PRMT1. It belongs to a family of signal transduction and activation of RNA (STAR) proteins which contain an extended K homology (KH) domain. Quaking (QKI) belongs also to the protein family of STAR proteins. The isoform one of QKI (QKI-5) was shown to be methylated. QKI in contrast to Sam68 seems not to be methylated by PRMT1 (Artzt and Wu, 2010; Cote et al., 2003). Here we identified QKI interacting with CARM1. PRMT8 itself is known to be methylated on arginine 58 and arginine 73 (Sayegh et al., 2007).

Lysine methylation of WIZ was found by literature search. WIZ contains several zinc fingers and was identified as a target site for methylation by EHMT2. The lysine methylation site on the WIZ protein functions as a methyl-specific binding site for HP1 (Rathert et al., 2008). WIZ

interacted with SUV39H1/2 and AOF2 (O'Carroll et al., 2000).

In addition, histone H2 is known to be methylated on three lysine residues (Beck et al., 2006). Histone H3 was also identified and is methylated on four arginine and five lysine residues by several PMTs (Jenuwein and Allis, 2001; Kouzarides, 2007; Martin and Zhang, 2005; Rea et al., 2000).

In summary, arginine methylation was 2.7 fold enriched in the interaction dataset. The absolute number of known arginine methylated proteins is small but we were able to identify eight arginine methylated non-histone proteins. The number of known lysine methylated non-histone proteins is even smaller still we were able to identify at least one non-histone protein known to be lysine methylated. The identification of several proteins known to be methylated emphasizes that methylation substrates can be detected by studying PMT- and PDeM-protein interactions. Quality of the data implied that the dataset is a resource for novel substrates.

Table 4: Identified known methylated proteins

GeneID	Symbol	Interacting bait	Modifying enzyme	K / R	Methylation function	Reference
3855	keratin 7 (**KRT7**)	AOF2	nd	R	nd	Ong 2004
2521	fused in sarcoma (**FUS**, ALS6, TLS)	SUV39H1	nd	R	nuclear-cytoplasmatic shuttling, involvement in ALS	Rappsilber 2003, Ong 2004, Tradewell 2011
3188	heterogeneous nuclear ribonucleoprotein H2 (H') (**HNRNPH2**)	PRMT6	nd	R	nd	by similarity Swissport
10492	synaptotagmin binding, cytoplasmic RNA interacting protein (**SYNCRIP**)	PRMT1, PRMT8	PRMT1	R	nuclear-cytoplasmatic shuttling	Passos 2006
2130	Ewing sarcoma breakpoint region 1 (**EWSR1**, EWS)	PRMT1	PRMT1, PRMT8	R	nuclear-cytoplasmatic shuttling , regulation of EWSR1 expression	Araya 2005, Kim 2008, Ong 2004
26135	SERPINE1 mRNA binding protein 1 (**SERBP1**, CGI-55)	PRMT8	PRMT1	R	nuclear-cytoplasmatic shuttling	Passos 2006
9444	QKI, KH domain containing, RNA binding (**QKI**)	CARM1	nd	R	nd	Côté 2003
56341	protein arginine methyltransferase 8 (**PRMT8**)	PRMT1, PRMT8	PRMT1, PRMT8	R	regulation of PRMT8 activity	Sayegh 2007
58525	widely interspaced zinc finger motifs (**WIZ**)	SUV39H1, SUV39H2	EHMT2	K	nd	Rathert 2008, Rathert 2009
8970	histone cluster 1, H2bj (**HIST1H2BJ**)	JMJD6	nd	K	nd	Beck 2006
440093	H3 histone, family 3C (**H3F3C**)	AOF2, JMJD2B, PRMT5, PRMT6, PRMT7, SETMAR, SMYD1, SMYD3, SUV39H1, WHSC1L1	AOF2, CARM1, PRMT5, PRMT6, SUV39H1, SUV39H2	R, K	regulation of chromatin structure	Rea 2000, Jenuwein 2001, Ong 2004, Martin 2005, Kouzarides 2007,

4.5.2 *Arginine methylation of PRMT interacting proteins containing RG-repeat motifs*

It is known that PRMT1, PRMT3, PRMT6 and PRMT8 preferentially methylate RG-rich regions (Lee and Stallcup, 2009). In a window of 12 amino acid RGs were searched and a RG-rich score calculated (Figure 20). Proteins having a RG-repeat score higher than 0.5 are shown in Figure 20. FUS and MYH7B had RG-rich regions but interacted with PKMTs and not with PRMTs. FUS, EWSR1, SYNCRIP and SERBP1 contained RG-rich regions and are known to be methylated by PRMT1 (Belyanskaya et al., 2001; Passos et al., 2006a; Passos et al., 2006b; Rappsilber et al., 2003). The SPSB2 protein was the only protein with a high RG-repeat score and interacting with PRMT1 but to our knowledge not known to be methylated. Therefore, it is a potential candidate for methylation. DOM3Z and ING5 interacted with PRMT6 and PRMT5, respectively. Both sequences contained RG-rich regions and thus are candidates for RG-methylation. The RS-rich related protein RXRC1 and U2AF1 (U2AF35) protein belong to the splicing factor SR family. Both proteins are involved in splicing and were identified interacting with JMJD6. The bromodomain containing protein (BRD4) and the ankyrin repeat domain 5 (ANKRD5) protein also interacted with JMJD6 and contained RG-rich sequence stretches.

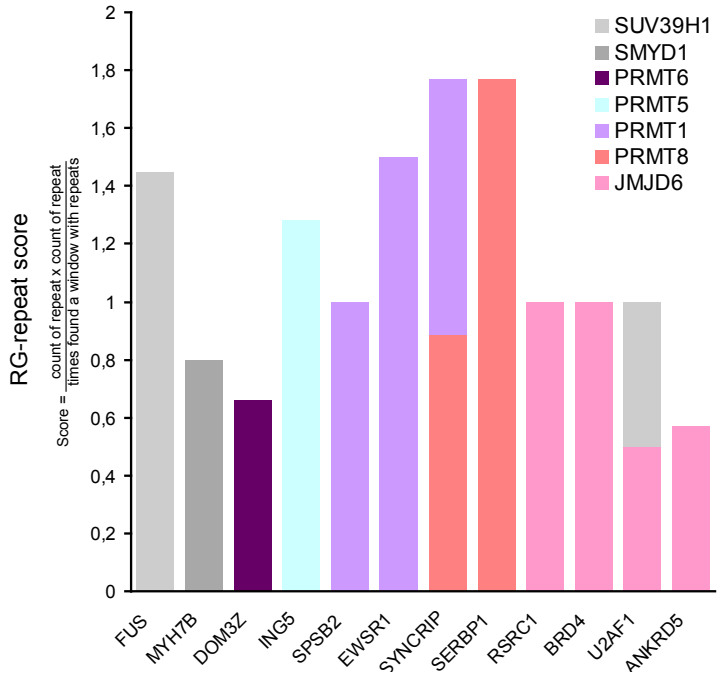

Figure 20: Proteins containing RG-repeats

Proteins having a RG-repeat score higher than 0.5 are shown. Colors indicate interacting enzymes. In a window of 12 amino acids RG-repeats were counted and a score was calculated as indicated.

In summary, RG-rich regions on proteins known to be methylated were verified. In addition, we identified RG-rich stretches on the PRMT interacting proteins DOM3Z, ING5 and SPSB2. Interestingly four proteins interacting with JMJD6 contained a RG-rich region.

4.5.3 *Enrichment of proteins involved in transcriptional regulation and DNA-binding*

in the network

We characterize the proteins identified in the screens for their molecular functions. Interacting prey proteins were analyzed and not the interactions because of the redundancy of the 22 baits containing enzymes involved in methylation. Enrichments in molecular functions were analyzed with the functional annotation tool called DAVID (http://david.abcc.ncifcrf.gov/home.jsp). GeneIDs of the identified prey proteins were uploaded as a multi-list file and examined against the complete set of proteins in the prey array.

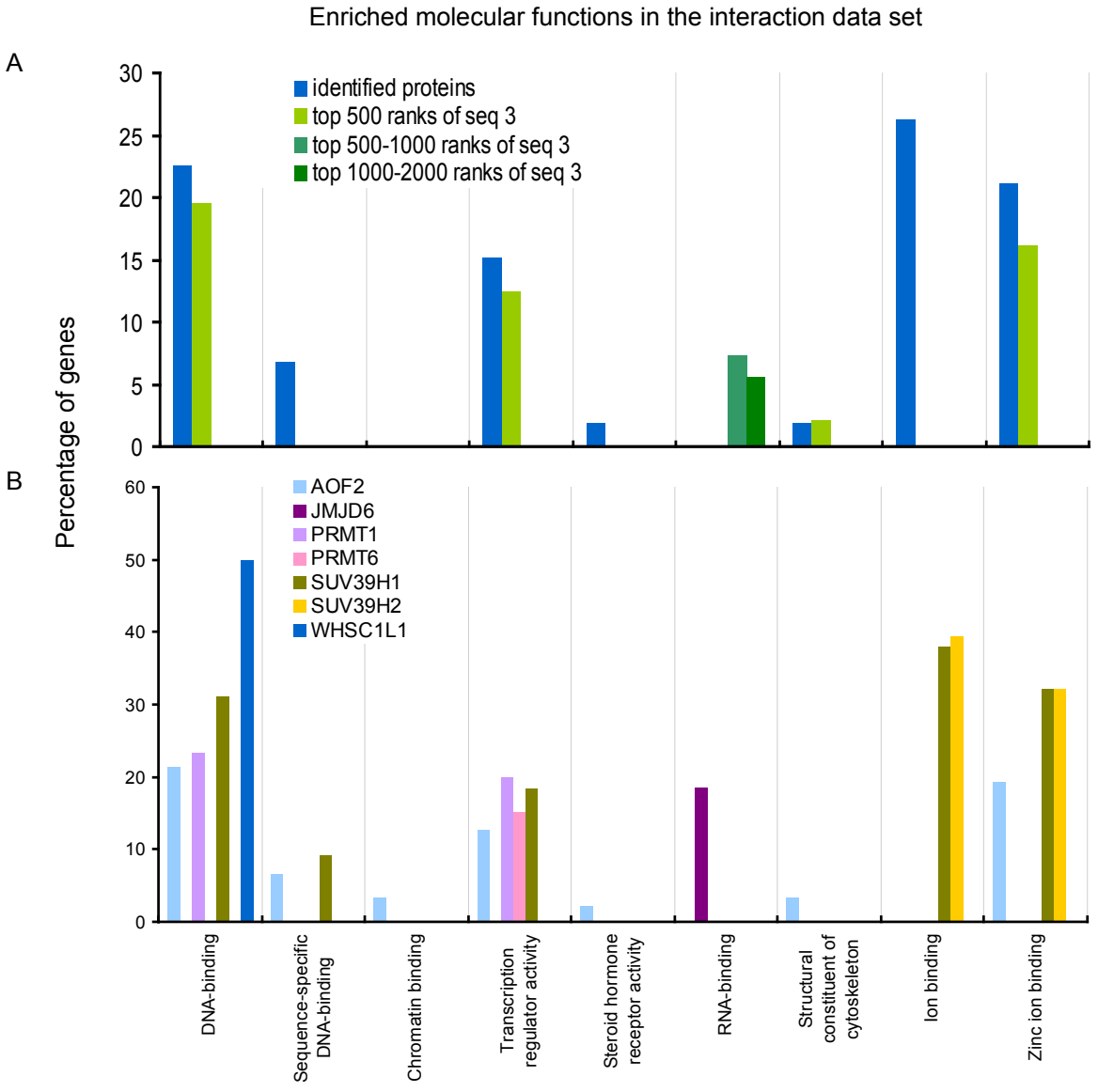

Figure 21: Molecular functions enriched in the PMT/PDeM interaction dataset

Identified proteins were assigned to molecular functions. Molecular functions enriched with a p-value lower than 0.05 are shown. (A) Enriched molecular function of all identified proteins (blue). Top 500 primary hits (light green), top 500-1000 primary hits (green) and top 1000-2000 primary hits (dark green) identified in the third Y2H-seq screen 3. (B) Enriched molecular function of proteins interacting with a specific enzyme. Gene ontology terms: DNA-binding GO:003677, sequence-specific DNA-binding GO:043565, chromatin binding GO:003682, transcription regulator activity GO:030528, steroid hormone receptor activity GO:03707, RNA binding GO:003723,

structural constituent of cytoskeleton GO:005200, ion binding GO:0432167 and zinc ion binding GO:008270.

In Figure 21A enriched molecular functions of PMT and PDeM interaction partners are shown in blue. Twenty-two percent of the proteins identified were enriched in DNA-binding, actually 8 % in sequence-specific DNA-binding. Additionally, zinc ion binding function was enriched with a p-value of 2.3E-7. Also transcription regulator activity was enriched which agrees with the enrichment of zinc ion binding function and DNA-binding function because those are often encoded by transcription factors. More specific steroid hormone receptor activity was enriched. Thus, there was a clear enrichment of transcription associated proteins interacting with PMTs. Two percent of the genes were involved in a distinct function: structural constituent of the cytoskeleton.

In a control analysis we examined if the enriched terms were seen in the background of obtained sequences. First the top 500 primary hits were tested including positively identified proteins, not tested (no clone available) and negatively tested proteins. The terms enriched in the positive identified proteins including DNA-binding, transcription regulator activity, structural constituent of the cytoskeleton and zinc ion binding were also enriched in the top 500 primary hits of the Y2H-seq screen 3. In the top 500 to 1000 and in the top 1000 to 2000 no enrichment was seen. This demonstrated the specificity of the enriched molecular functions in the top 500 genes. Only RNA binding was enriched for higher ranks what was not true for the positively identified proteins and for the top 500 proteins.

In order to yield a more specific picture of the interactions all positive identified proteins were analyzed with respect to their interacting PMTs and PDeMs (Figure 21B). SUV39H1/2 and WHSC1L1 are histone methyltransferases and regulate epigenetic transcription (Kim et al., 2006b; Schotta et al., 2003). We identified 50 % of the genes interacting with WHSC1L1 to bind DNA. Proteins interacting with SUV39H1 were also enriched for DNA-binding and additional in transcriptional regulation activity. Those proteins contained zinc fingers and thus 32 % of the SUV39H1 interacting proteins were enriched in zinc ion binding function.

AOF2 is mainly known as histone demethylase and involved in transcriptional regulation (Klose and Zhang, 2007; Shi et al., 2004). P53, MYPT1 and DNMT1 are the only non-histone substrates known of AOF2 (Cho et al., 2011; Huang et al., 2007; Wang et al., 2009). Here, AOF2 was enriched in several molecular functions including sequence-specific DNA-binding and zinc ion binding. Those functions play a role in transcriptional regulator activity which was also enriched in AOF2 interacting proteins. Enrichment in structural constituent of cytoskeleton was specific to AOF2 and attributed to the keratin genes interacting with AOF2.

PRMTs methylated non-histone substrates and are involved in several cellular processes like RNA processing, signal transduction and DNA repair (Bedford and Richard, 2005). PRMT1 and PRMT6 interacting proteins showed enrichment in transcription regulator activity and PRMT1 interacting protein additionally in DNA-binding.

Interestingly, JMJD6 interacted with RNA binding proteins even though they were not enriched in the set of positively identified interactions as such. The identified RNA binding proteins were associated with spliceosomal function which linked JMJD6 to a role in splicing. Therefore, it is important to analyze.

4.5.4 Enrichment of specific domains in the PMT- and PDeM interaction dataset

Interaction as well as function is mediated by protein domains in higher eukaryotes. Therefore, we analyzed the set of interacting proteins with the annotation online tool DAVID (as described before in 4.5.3). In Figure 22 specific domain occurrence is shown for proteins interacting with all enzymes (Figure 22A), functional related enzymes (Figure 22B) and specific enzymes (Figure 22C).

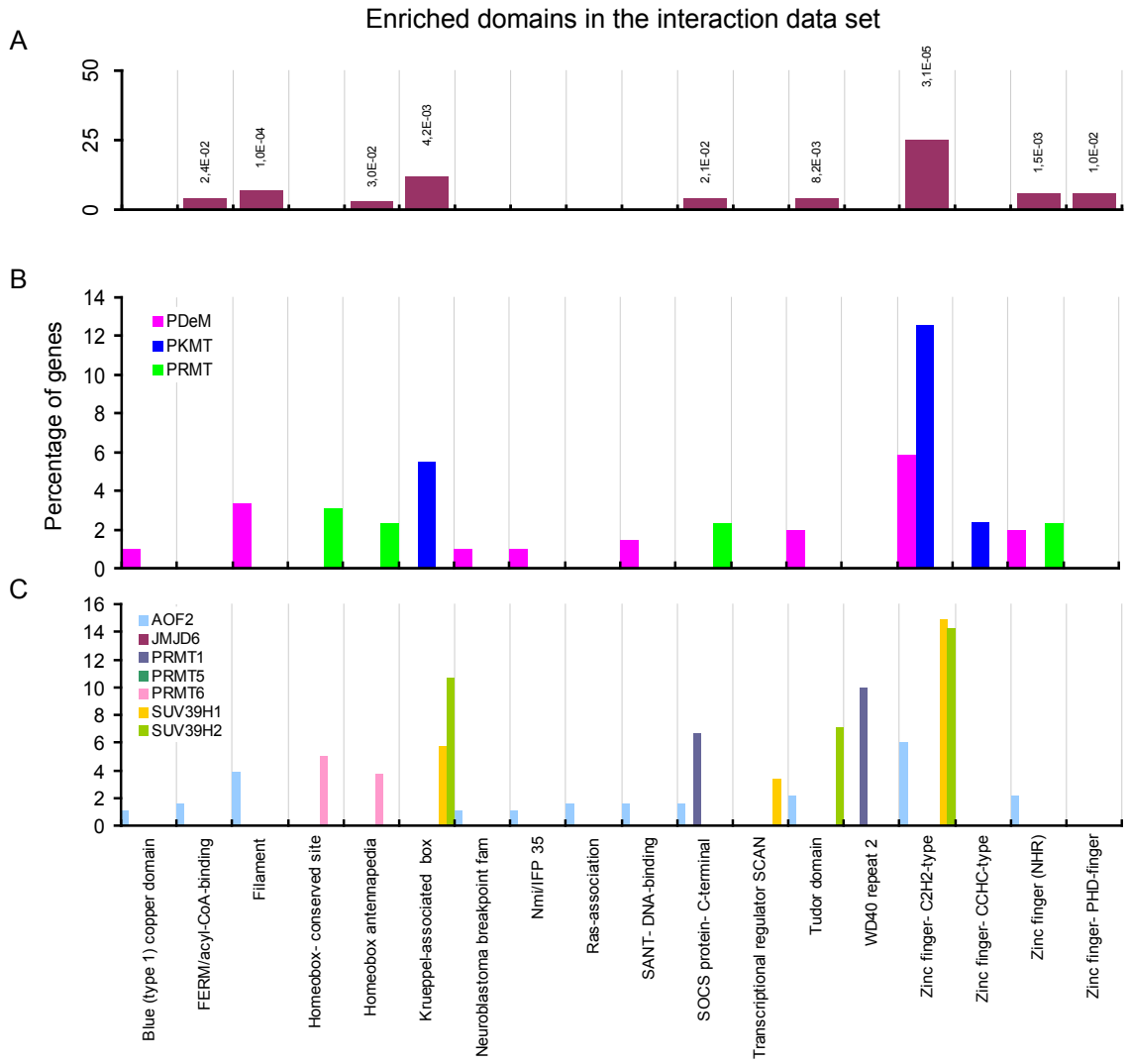

Figure 22: Domains enriched in the PMT/PDeM interaction dataset

Domains enriched with a p-value lower than 0.05 are shown. (A) Domains enriched in all interacting

proteins. (B) Domains enriched in proteins interacting with PDeMs, PKMTs or PDeMs. (C) Enriched domains in proteins interacting with specific enzymes. Gene ontology terms: Blue (type 1) copper domain IPR000923, FERM/acyl-CoA-binding IPR014352, Filament IPR016044, Homeobox-conserved site IPR017970, Homeobox antennapedia IPR001827, Krueppel-associated box IPR001909, Neuroblastoma breakpoint fam IPR010630, Nmi/IFP 35 IPR009909, Ras-association IPR000159, SANT-DNA-binding IPR001005, S/T protein kinase-related IPR017442, SOCS protein-C-terminal IPR001496, Transcriptional regulator SCAN IPR003309, Tudor domain IPR002999, WD40 repeat 2 IPR019782, Zinc finger-C2H2-type IPR007087, Zinc finger-CCHC-type IPR001878, Zinc finger (NHR) IPR001628 and Zinc finger-PHD-finger IPR019787.

The conserved homeobox domain was enriched in the complete set of interacting proteins (Figure 22A), further in proteins interacting with PRMTs (Figure 22B) and specifically in proteins interacting with PRMT6 (Figure 22C). SUV39H1 and SUV39H2 interacted with proteins containing the krueppel-associated box and the zinc finger C2H2-type domain. The Krueppel-associated box is commonly found in C2H2-type zinc finger proteins (Urrutia, 2003). SUV39H1 also interacted with the transcriptional regulator domain SCAN known as a highly conserved domain motif found in the C2H2-type zinc finger proteins (Williams et al., 1999). The C2H2-type zinc finger motif is present in several transcription factors (Krishna et al., 2003). Hence, proteins interacting with SUV39H1 and SUV39H2 contained domains involved in transcriptional regulation which concurred with the enrichment in transcriptional regulator activity of protein interacting with SUV39H1 (4.5.3). AOF2 interacted with a significant number of keratins containing the filament domain. This resembles the enriched molecular function in structural constituent of cytoskeleton of AOF2 interacting proteins (4.5.3). Interestingly SOCS box and tudor domain containing proteins were enriched in proteins interacting with PRMT1. SOCS proteins are inhibitors of cytokine signaling pathways and regulate immune-cell function (Yoshimura et al., 2007). Tudor domains bind methylated residues (Taverna et al., 2007).

In summary, we identified six domains which were enriched in the complete set of interacting proteins and in proteins interacting with a specific enzyme or a related group of enzymes. In addition, we identified WD40 repeat 2 domain containing proteins enriched in proteins interacting selectively with PRMT1.

4.6 The PMT and PDeM interactome network

To get an overview about interactions of the different PMTs and PDeMs a network was drawn containing the complete interaction dataset. The molecular function, the cellular component and the domain annotation of the prey proteins were integrated in the network to visualize the results of the previous sections. The interaction dataset was imported into cytoscape an open source bioinformatics software platform for visualizing molecular interaction networks. The illustrated network contains 523 PPI between 324 prey proteins and 22 bait proteins identified in the four screens (Figure 23). The 22 baits were drawn as pink squares. The color of the nodes indicated the molecular function, the shape of the nodes indicated the cellular component and domains were

indicated by the border color. Green, round nodes without a colored border had no annotation for molecular function, cellular component and domain annotation in the network.

AOF2 had the largest set of 182 interacting prey proteins. Also SUV39H1 and PRMT6 had many interactions. A representative interaction dataset was obtained for PRMT1, PRMT5, PRMT8, JMJD6, SUV39H2, SMYD1 and WHSC1L1. Some prey proteins overlapped between the different baits for example SUV39H1 and AOF2 shared some of their interaction partners but the main part of the interacting prey proteins was specific to AOF2.

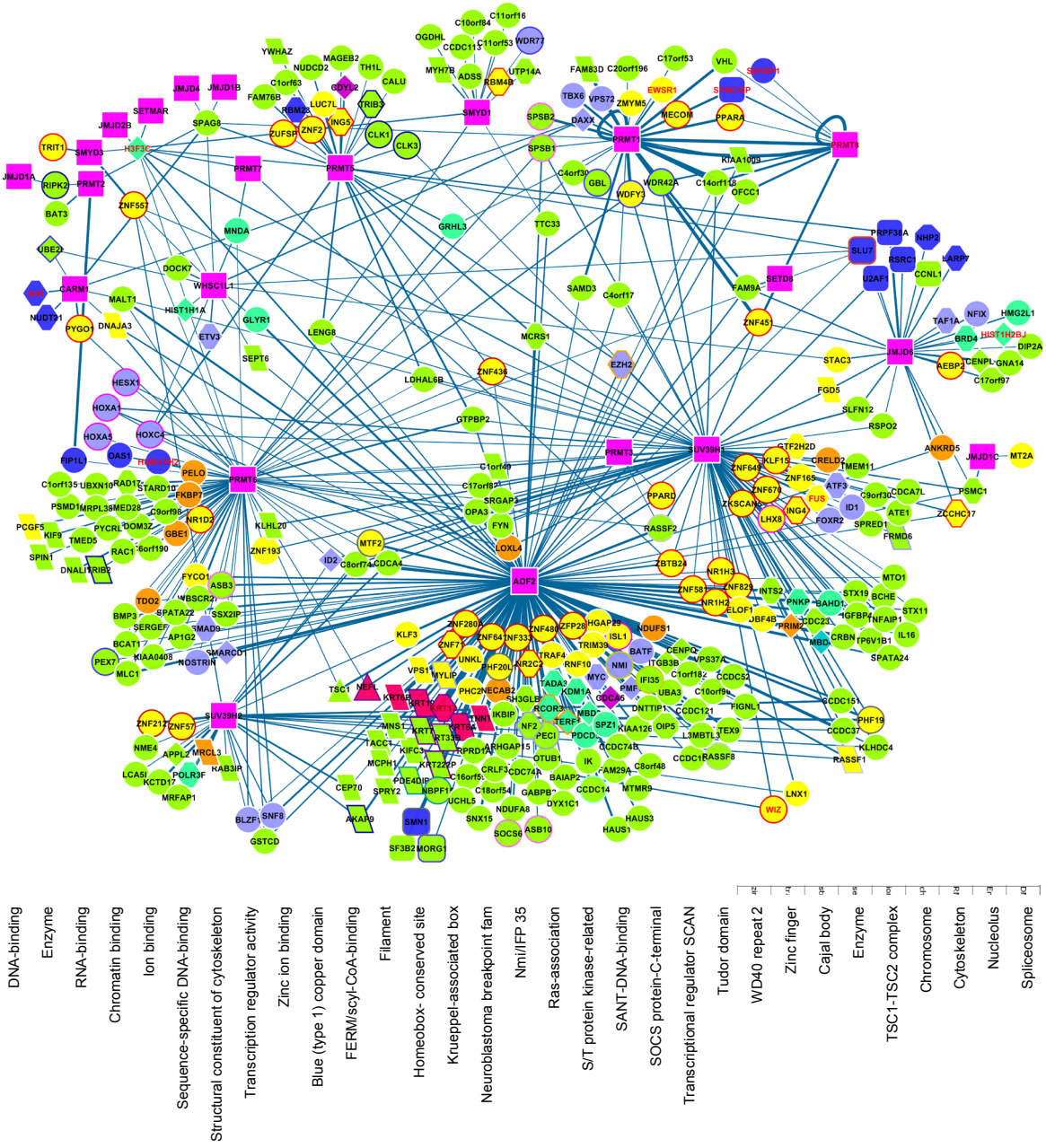

Figure 23: PMT and PDeM interaction network obtained in the Y2H screens

Nodes represent proteins and lines represent protein interactions. The line width indicates how often this interaction was found using different replicas or different clones for one protein. Baits containing the enzymes are indicated as pink squares. All other nodes show interacting prey proteins and are labeled as followed: Node shape indicates the cellular component of the protein, the node color

indicates the molecular function of the protein and node boarder color indicates domain of the protein. Labels of known methylated proteins are red. Nodes are labeled with the official entrez gene symbol. Additional information about the molecular function, cellular component and domain annotation of the proteins were gained from DAVID and integrated in the network.

SMYD1 only shared three interaction partners with other enzymes pointing out the high interaction specificity of SMYD1. At large, most of the baits interacted very specific and shared only a small part of there interacting proteins with other baits confirming the high substrate specificity of the enzymes seen in Figure 18. Many AOF2 interacting proteins had a zinc ion binding function (yellow node) and some of those had a zinc finger domain (red boarder). Proteins having a DNA-binding function often had also a zinc ion binding function. In the network the zinc ion binding was preferentially illustrated. Many of the AOF2 interacting proteins were components of the cytoskeleton (parallelogram). Some of those had a purple boarder color indicating filament domain and some were red indicating a function in structural constituent of the cytoskeleton. SUV39H1 interacted with many ion binding proteins (yellow node) and zinc finger domain containing proteins (red boarder). The proteins interacting with SUV39H1 were localized in the nucleus (hexagon) and at the chromosome (diamond). JMJD6 interacting proteins were often localized in the nucleus or at the spliceosome (round rectangle) and bind RNA (blue node). PRMT8 and PRMT1 had a large overlap of interacting proteins (further described section 4.6.1 below). PRMT5 interacted with three S/T kinases (blue boarder). CARM1 and WHSC1L1 had a diverse set of interacting proteins. Noticeable were the transcriptional regulator proteins (purple node) with the homeobox domain (pink boarder) interacting with PRMT6.

Molecular function and protein domains enriched in the protein interaction dataset are summarized in the interaction network which presents first proteome-wide dataset of PMTs and PDeMs interacting proteins and potential methylated proteins, revealing common as well as distinct partners and functions.

4.6.1 Subnetwork of common interaction partners between PRMT1 and PRMT8

We suggested that some of the PRMT1 interacting proteins were methylated and wanted to visualize the most prominent candidates to be methylated in a subnetwork of PRMT1 and PRMT8. Figure 24 shows the proteins interacting with PRMT1 and PRMT8.

PRMT1 and PRMT8 form homodimer and heterodimer. A part from that, they share relatively many interaction partners. Especially, PRMT8 interactors were all shared with PRMT1 except SERBP1 interacted only with PRMT8. However, SERBP1 is known to be methylated by PRMT1 (Passos et al., 2006a). PRMT1 and PRMT8 clustered in the network because of the high overlap of interacting proteins. Probably because PRMT8 amino acids are to 80 % identical to PRMT1 (Kim et al., 2008). DAXX, TBX6 and VPS72 had a transcriptional regulatory activity (purple). FAM83D and KIAA1009 were localized at the cytoskeleton (parallelogram).

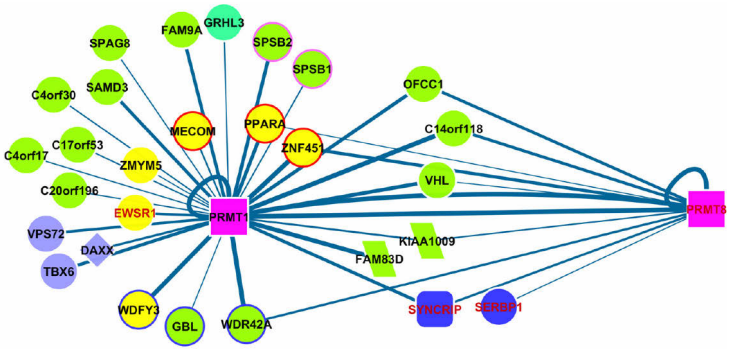

Figure 24: Subnetwork of PRMT8 and PRMT1

Interactions of PRMT8 and PRMT1 are shown. For labels see legend of Figure 23.

SYNCRIP and SERBP1 were RNA binding proteins (blue node color). SYNCRIP was localized at the spliceosomal complex (round rect). PRMT8, SYNCRIP, EWSR1 and SERBP1 are known to be methylated indicate by a red node label color. SYNCRIP, EWSR1 and SERBP1 were methylated in the RG-rich regions (Belyanskaya et al., 2001; Passos et al., 2006a; Passos et al., 2006b; Sayegh et al., 2007). SPSB2 also contained a RG-rich region and is presumable methylated (Figure 20). SPSB2 and the closely related protein SPSB1 also interacted with PRMT1. Some of the RG-rich regions were also present in SPSB1 and might be methylated, too (see discussion). Other proteins like the WDR42A, GBL and WDFY3 contained WD repeat domains (blue border color) which were enriched in the set of PRMT1 interactions (see domain enrichment analysis in 4.5.3). The PRMT1/PRMT8 subnetwork contained five non characterized proteins only named after their chromosomal localization. Also the OFCC1 is a protein with unknown function.

4.7 Identification of methylated residues in PMT interacting proteins

4.7.1 Detection of methylated proteins in a methylation assay using radioactive AdoMet

To establish exemplarily that proteins in the PPI dataset are methylation substrates we assessed the methylation of candidate proteins. Detection of methylation is difficult because no affinity reagents or high-quality antibodies are available (Komyod et al., 2005; Levy et al., 2011; Pahlich et al., 2006). Hence, methylation is mostly discovered by radioactive labeling or mass spectrometry. Here, radioactive labeled methyl donor ^{3}H *S*-adenosylmethionine (^{3}H SAM, ^{3}H AdoMet) was used in an *in vitro* methylation assay to detect methylation (Figure 25A). The candidate proteins were expressed in *E. coli* to yield sufficient amounts of the proteins to be methylated. This has the advantage that proteins expressed in *E. coli* are in general not posttranslational modified. The disadvantage was that not all human proteins tested were expressed in *E. coli*. It is known that eukaryotic proteins are often difficult to express in *E. coli* in soluble from (Gilbert and Albala, 2002). Only a certain fraction of the proteins is overproduced in *E. coli* in sufficient yield without formation of inclusion body aggregates or proteolytic degradation (Bussow et al., 2005). Candidate

proteins potentially methylated were shuttled into an *E. coli* expression vector in which the ORF is fused to a hexahistidine-tag (his-tag). Then the *E. coli* expression strain SCS1 was transformed with the expression vector. Bacteria were grown until the exponential phase was reached and then protein expression was induced by addition of IPTG. The cleared *E. coli* lysate was transferred into a microfuge tube containing nickel agarose beads which bind his-tagged proteins. Unbound cell material was washed off. The immobilized candidate proteins are shown on a SDS gel in Figure 25D. In parallel, enzymes identified interacting with the candidate proteins were overexpressed in mammalian HEK293 cells. The HEK293 mammalian cell lysate containing the enzyme was mixed with radioactive labeled methyldonor (^{3}H-AdoMet, Figure 25B) and added to the candidate protein immobilized on nickel agarose beads. As control PBS buffer mixed with ^{3}H AdoMet was added to the candidate protein immobilized on nickel agarose beads. After allowing the methylation to occur the unbound radioactive labeled material was washed of by intensive washing. The nickel agarose beads with the bound his-tagged candidate proteins were measured in a scintillation counter to detect incorporated radioactive labeled material. Figure 25C reports the incorporation of radioactive methyl to several reactions with different proteins. SYNCRIP incubated with HEK cell lysate overexpressing PRMT1, but not when in control incubation, showed a high ^{3}H signal indicating methylation of SYNCRIP. Less but still significant was the incorporation of ^{3}H methyl in the positive control PABP1. ATE1 and SNX15 as well as the negative controls which were incubated with buffer instead of HEK cell lysate showed no incorporation of ^{3}H methyl. PSMC1 showed a weak signal compared to SYNCRIP which nevertheless might indicate methylation.

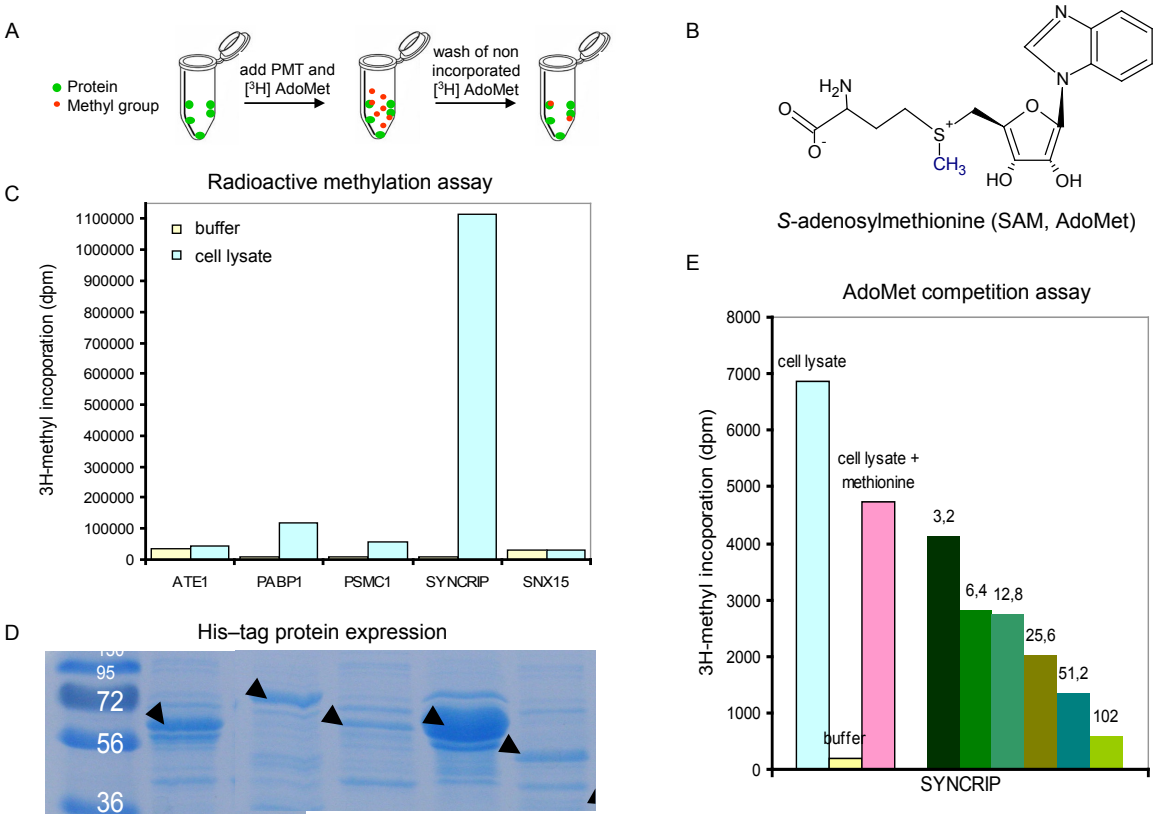

Figure 25: Methyltransferase assay using ^{3}H AdoMet

(A) Schematic representation of the radioactive labeled methyltransferase assay. (B) Structural formula of S-adenosylmethionine (AdoMet, SAM). (C) Incorporation of ^{3}H methyl groups into his-tag purified candidate proteins. (D) Expression control of purified his-tagged proteins (E) Competition of radioactive ^{3}H AdoMet with increasing amounts of non radio labeled (^{1}H) AdoMet using SYNCRIP. Positive control (light blue) was incubated with HEK cell lysate and ^{3}H AdoMet. Negative controls were incubated with buffer and ^{3}H AdoMet (yellow) or HEK cell lysate, ^{3}H AdoMet and methionine (pink). Six samples (green nuance) were incubated with HEK cell lysate, ^{3}H AdoMet and increasing amounts of ^{1}H AdoMet indicated above the bars.

To approve that the incorporated ^{3}H methyl was due to methylation of the candidate proteins we tested if the incorporation was dependent on AdoMet. Therefore, we tested if non radioactive labeled AdoMet (^{1}H AdoMet) was able to compete for radioactive ^{3}H AdoMet. SYNCRIP was used as candidate protein in the methylation assay which was performed as described above. Increasing amounts of ^{1}H AdoMet in the HEK 293 cell lysate competed for ^{3}H AdoMet (Figure 25E) in a dose dependent manner demonstrating that the ^{3}H methyl incorporation was due to AdoMet and hence enzymatic activity. In summary, we established an *in vitro* methylation assay detecting incorporation of a radioactive labeled methyl groups in purified candidate proteins using scintillation counting.

4.7.2 Mapping methylation sites using in vitro methylation coupled to mass spectrometry

We detected methylation on proteins in a radioactive methylation assay. To discover methylation sites mass spectrometry (MS) was used. The *in vitro* methylation assay was carried out as described above but using non radioactive AdoMet. Briefly, eight proteins including, PYGO1, QKI, SERBP1, SAMD3, OFCC1, WDR42A and SYNCRIP identified interacting with PRMT1 or CARM1 were expressed in *E. coli* and immobilized on beads. Protein A tagged PRMT1 and CARM1 were transfected and expressed in mammalian cells. The immobilized proteins were incubated with the mammalian cell lysate expressing the appropriate enzyme. After washing the protein was applied on a SDS-PAGE. The band containing the protein of interested was cut out and digested with trypsin. Trypsin cleaves peptide chains mainly at the carboxyl side of arginines and lysines. As SYNCRIP contains many arginines it was partially digested to get suitable peptides for the MS analysis. The peptide fragments were separated by liquid chromatography and then analyzed by tandem mass spectrometry (LC-MS/MS) (by David Meierhofer, OWL Sauer, MPIMG in Berlin). Data were analyzed by MaxQuant (http://maxquant.org/) the quantitative proteomics software package to analyze large mass-spectrometric datasets. We scanned for mono- and dimethylarginines as well as for mono-, di- and trimethyllysines. Spectra of methylated and unmethylated peptides were manually inspected. Table 5 summarizes the results listing the tested proteins and reporting non-methylated and methylated peptides identified. We could identify methylation sites on seven of the nine tested proteins.

SYNCRIP a known PRMT1 substrate was identified with sequence coverage of 85 % and 15 methylated arginine residues. Seven arginines were identified to be monomethylated, three

dimethylated and five mono- and dimethylated. Five arginines in the three RRM domains of SYNCRIP were monomethylated. The arginine 409 was mono- and dimethylated. In Figure 26 the fragment spectrum of unmethylated, mono- and dimethylated peptide containing the R409 is shown. The arginine 409 was mono- and dimethylated seen by the shift of the y-ionen. A methyl group has a mass of 14 Da. Hence, the shift of a demethylated peptide fragment to a methylated was 14 Da if the peptide fragment was charged twice the shift was only 7 Da and so forth. For example, the mass of the y_{14}^{2+} ion increased by 7 Da from spectrum in Figure 26A to B and further 7 Da in spectrum C indicating mono- and dimethylation of the arginine, respectively.

Table 5: Identified methylated peptides

Accession no.	Protein name	Total peptide	Unique peptide	Sequence coverage	Site	Methylated peptide
Q5JZB8	Spindlin family, member 2B (**SPIN2B**)	600	42	84%		
					R15	TPNAQEAEGQQTR(me)
					R53	NIVGCR(me)
						RNIVGCR(me)
					R100	YDGIDCVYGLELHR(me)
						YDGIDCVYGLELHR(me)DER
					R200	IMPESSESPPTER(me)
Q9Y3Y4	Pygopus homolog 1 (**PYGO1**)	3	3	9%		
Q96EY1-2	DnaJ homolog subfamily A member 3, mitochondrial (**DNAJA3**)	64	34	66%		
					R58	KLSVPAFASSLTSCGPR(me)
						LSVPAFASSLTSCGPR(me)
					R293	GSIIISPCVVCR(me)
Q96PU8	Protein quaking (**QKI**)	336	52	64%		
					R43	LMSSLPNFCGIFNHLER(me)
						KLMSSLPNFCGIFNHLER(me)
					K177	LLVPAAEGEDSLKmeK
					R242	IITGPAPVLPPAALR(me)TPTPAGPTIMPLIR
						IITGPAPVLPPAALR(dime)TPTPAGPTIMPLIR
Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein (**SERBP1**)	135	34	49%		
Q8N6K7	Sterile alpha motif domain-containing protein 3 (**SAMD3**)	269	29	27%		
					K157	SYVLPEFPYDVK(me)
Q8IZS5	Orofacial cleft 1 candidate gene 1 protein (**OFCC1**)	222	46	80%		
					K57	EATEGTGNPAFNMSSPDLSACQTAEK(me)
						EATEGTGNPAFNMSSPDLSACQTAEK(me)K
						EDREATEGTGNPAFNMSSPDLSACQTAEK(me)
						EDREATEGTGNPAFNMSSPDLSACQTAEK(me)K
					R105	NYFDPLMDEEINPR(me)
					R113	QCATEVSR(me)
						QCATEVSR(me)EDDDR
						QCATEVSR(me)EDDDRIFYNR
					R195	TIILLGSSPLEQEIR(me)
Q5TAQ9-2	DDB1- and CUL4-associated factor 8 (DCAF8, **WDR42A**)	823	41	86%		
					R33	TDLANGSLSSSPEEMSGAEEGR(me)
					R94	GTDTESSGEDKDSDSMEDTGHYSINDENR(me)
					R115	SEEEEEEEEEEEEEQPR(me)
						VHDRSEEEEEEEEEEEEEQPR(me)
					R134	DQDSSDDER(me)
					R177	FVYEACGAR(me)
					R204	LQHGLEGHTGCVNTLHFNQR(me)
					R252	QGSIIATER(me)
O60506-4	Heterogeneous nuclear ribonucleoprotein Q (**SYNCRIP**)	634	121	85%		
					K100	SAFLCGVMK(me)
					K213	GYAFVTFCTK(me)
						LMMDPLTGLNRGYAFVTFCTK(me)
					R229	LYNNHEIR(me)
					R232	SGK(me)HIGVCISVANNR
					R244	HIGVCISVANNR(me)
					R286	NR(me)GFCFLEYEDHKTAAQVK
					K321	NLANTVTEEILEK(me)
					R346	LKDYAFIHFDER(me)
					R409	NQMYDDYYYYGPPHMPPPTR(me)GR
						NQMYDDYYYYGPPHMPPPTR(di)GR
						QAAKNQMYDDYYYYGPPHMPPPTR(me)GR
						QAAKNQMYDDYYYYGPPHMPPPTR(di)GR
					R461, R463	GGYEDPYYGYEDFQVGAR(me)GR(me)
					R466, R469, R475	GGR(di)GAR(di)GAAPSR(me)
					R475	GAAPSR(di)GR
					R483	GAAPPR(di)GR
					R491	AGYSQR(me)GGPGSAR
						AGYSQR(di)GGPGSAR
					R491, R498	AGYSQR(di)GGPGSAR(me)GVR
						AGYSQR(di)GGPGSAR(di)GVR
					R504	GAR(me)GGAQQQR
						GAR(di)GGAQQQR
						GVRGAR(di)GGAQQQR

The mass of the b_{11}^{+} -NH3 ion did not change because the methylated arginine was not contained in the b_{11} peptide fragment. The b_{25} ion contained the methylated arginine 409 but not the arginine 411. The mass change of the b_{25} ion pointed to the methylation of arginine 409. The mass of the b_{25}^{4+} ionen increased by around 7 Da (3.5 Da methyl plus 4 Da ammoniac) in the spectrum (Figure 26A to B). The b_{25}^{3+} peptide fragment was around 4.7 Da bigger as the calculated mass without two methyl groups. This indicated mono- and dimethylation of R409. Nine methylation sites were identified in the RG-rich (448 to 559) domain of SYNCRIP most of them dimethylated. Additionally, three lysines were identified to be monomethylated.

SERBP1, known to be methylated by PRMT1 (Passos et al., 2006a), was identified by mass spectrometry with sequence coverage of 49 % but no methylation was identified. Methylation was expected to be on RG-rich regions but only one peptide containing a RG-rich region was identified in the MS spectrum.

SAMD3, OFCC1 and WDR42A were identified interacting with PRMT1 and tested for methylation. We identified SAMD3 with sequence coverage of 27 % and one peptide methylated on lysine 157. OFCC1 was identified with 46 unique peptides. The lysine 57 was methylated and identified by four unique methylated peptides. Three other arginine sites were monomethylated. None of them was in a RG-rich region. For the WDR42A protein 41 unique peptides were identified with sequence coverage of 86 %. Seven arginines were monomethylated whereof three occurred in a RG-rich region. The R134 was monomethylated (Figure 27). The 14 Da shift of the y_{7+} ion matches the size of one methyl group (compare Figure 27A to B). The b_{2+} ion instead had the identical *m/z* in Figure 27A and B because the methylated arginine was not contained in the b_{2+} peptide fragment.

We also identified SPIN2B to be methylated. SPIN2B is a spindilin family member as the SPIN1 protein which was identified interacting with PRMT6. We tested SPIN2B in the methylation assay and identified four methylated arginines.

We analyzed also the CARM1 interactors PYGO1, DNAJA3 and QKI in the LC-MS/MS approach. Only three peptides were identified for PYGO1 resulting in a sequence coverage of 9 %. None of the three identified peptides was methylated. DNAJA3 was identified with sequence coverage of 66 % and two methylated arginines. The protein QKI was identified with 52 unique peptides. QKI is known to be methylated but the sites of methylation and the methylating enzyme not (Cote et al., 2003). We identified QKI interacting with CARM1 in the Y2H screens. In the methylation assay coupled to LC-MS/MS arginine 43 and 242 were identified to be methylated. In addition, one methylated lysine was discovered.

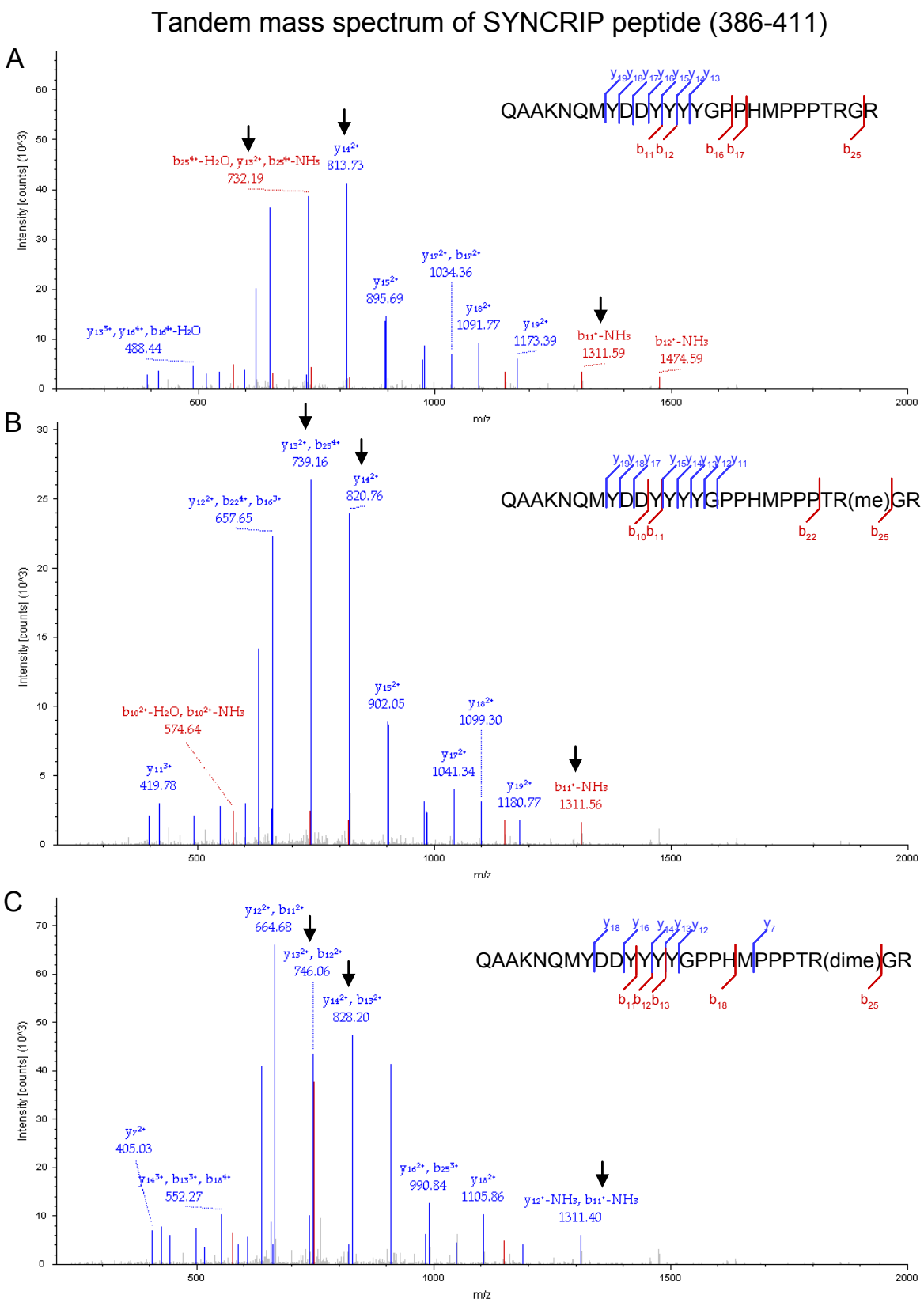

Figure 26: Tandem mass spectrum of the tryptic peptide 386-411 of SYNCRIP protein

(A) Tandem mass spectrum of quarto charged ion at *m/z* 779.9 corresponding to tryptic peptide QAAKNQMYDDYYYYGPPHMPPPTRGR. (B) Tandem mass spectrum of quarto charged ion at *m/z* 783.4 corresponding to tryptic peptide QAAKNQMYDDYYYYGPPHMPPPTR(me)GR. (C) Tandem mass spectrum of quarto charged ion at *m/z* 786.9 corresponding to tryptic peptide QAAKNQMYDDYYYYGPPHMPPPTR(dime)GR.

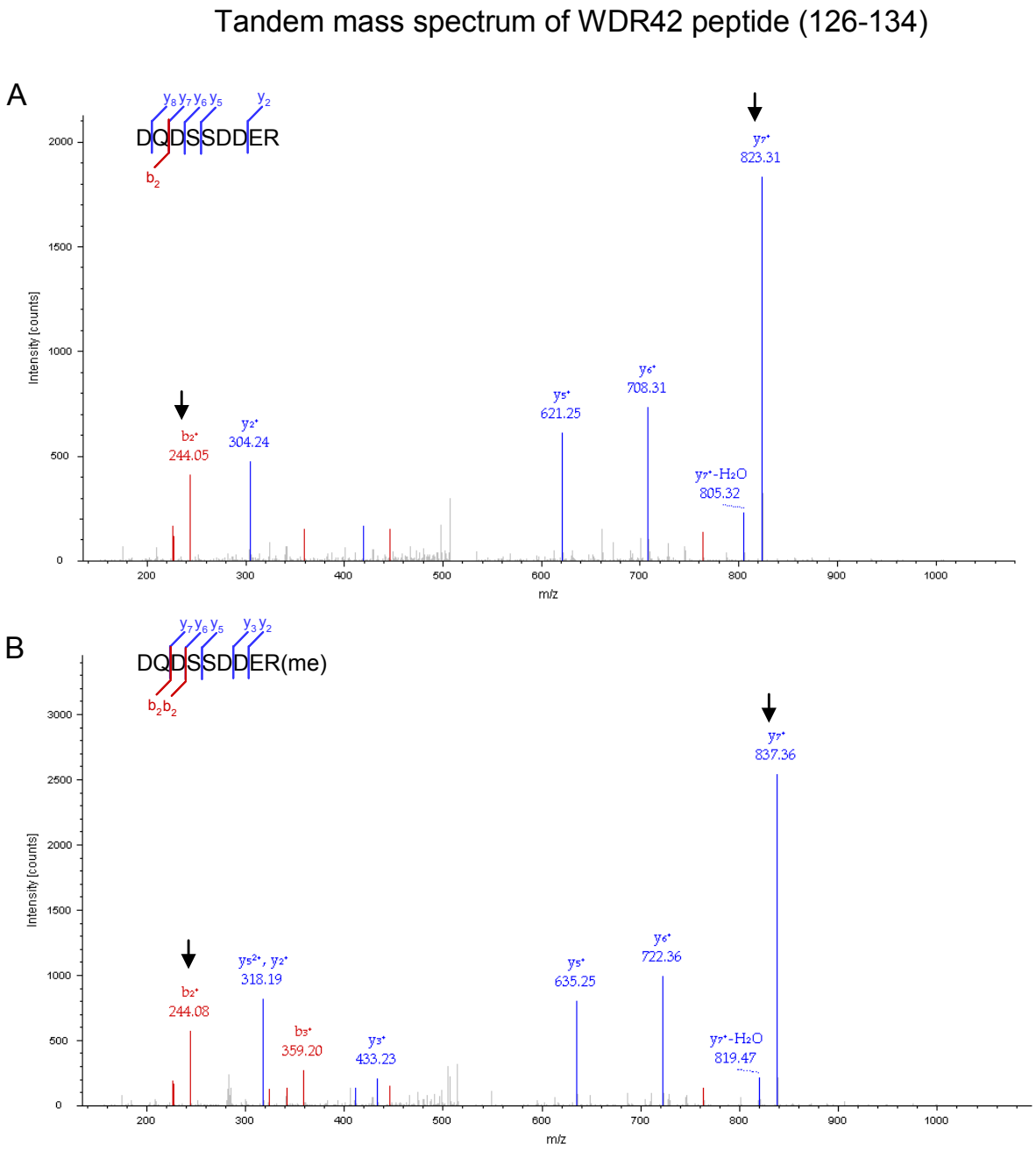

Figure 27: Tandem mass spectrum of the tryptic peptide 126-134 of WDR42A protein

(A) Tandem mass spectrum of double charged ion at *m/z* 533.7 corresponding to tryptic peptide DQDSSDDER. (B) Tandem mass spectrum of double charged ion at *m/z* 540.7 corresponding to tryptic peptide DQDSSDDER(me).

In summary, we identified seven proteins to be methylated on various arginine and lysine sites. For the known methylated proteins QKI and SYNCRIP we identified the sites of methylation. SERBP1 RG-rich peptides were not discovered and therefore probably no methylation of SERBP1. For the PRMT1 interacting WDR42A and OFCC1 we identified new arginine methylation sites. SAMD3 was methylated on a lysine residue. For the CARM1 interacting PYGO1 protein no methylation was detected probably because of the low sequence coverage. DNAJ3 was identified to have two arginine methylation sites.

5 Discussion

5.1 Quality estimates of Y2H protein interaction data

The Y2H system is a widely-used tool for the discovery of protein-protein interactions (PPIs). At present, it is one of the most powerful methods for the generation of proteome-wide, binary PPI maps and will play a crucial role in the whole-organism interactome mapping (Schwartz et al., 2009; Venkatesan et al., 2009; Worseck et al., 2012; Yu et al., 2011). The "precision" of PPI data can be determined as the fraction of observed pairs in an interactome dataset that are true interactions. If a positive reference set and high-throughput Y2H (HT-Y2H) dataset are retested in a third assay, the mammalian protein-protein interaction trap assay (MAPPIT) (Eyckerman et al., 2001), the precision is comparable. Hence, existing human HT-Y2H data have a high precision (Venkatesan et al., 2009). However, the same study shows that the data have relatively low coverage (5-20 %), determined as the fraction of interactions found over the known interactions in the search space. Therefore, Braun et al. suggested relaxing screen stringency thereby increasing the number of interactions that will be detected (Braun et al., 2009). This also increases the number of false positives (non-interacting protein pair reported as interacting) and hence decreases the precision of the data obtained. The authors suggest reporting a p-value for all published interactions, which will be derived from assessing the interactions in additional assays, and thus ensuring their high quality. However, the screening data produced would have a lower quality and the workload in a secondary assay would be high. In contrast, Schwartz et al. presented an alternative model to efficiently detect proteome-wide interactions suggesting to use a large set of stringent assays, to always test those protein pairs for which not enough evidence has been collected to decide whether they interact or not (Schwartz et al., 2009). We designed our new method in line with the latter proposal. We wanted to keep the primary screening stringency and increase the number of interactions by increasing the sensitivity. There are several factors affecting the sensitivity. One is the "assay sensitivity" which is determined as the subset of interactions a method detects. Hence, the coverage of PPI data can be increased by combining different PPI detection methods (Schwartz et al., 2009) or different versions of the same assay (Sanderson, 2009; Stelzl and Wanker, 2006; Suter et al., 2008). Provided that different Y2H systems are producing high precision data, the parallel use of several different Y2H setups would simply increase sensitivity. This would increase the workload, though (Figeys, 2008; Rajagopala et al., 2009; Schwartz et al., 2009). Another factor affecting the sensitivity is the "sampling sensitivity" which is determined as the fraction of all identifiable interactions that are found in a single trial of an assay performed. The sampling sensitivity of Y2H-matrix per screen is about 45 % and at least six screens are needed to reach 90 % saturation in our system resulting in six times the workload (Venkatesan et al., 2009).

In summary, the stringent Y2H approach detects high-quality PPIs but due to the low

sensitivity the data coverage is low. The sensitivity can be increased by repeating the screens or combining different assays but is laborious. Therefore, we developed a Y2H approach with a reduced workload.

5.1.1 A new Y2H-seq approach reducing the workload and increasing the sensitivity

To reduce the workload of a Y2H screen we decided to pool the preys and identify interacting prey proteins by second generation sequencing in a new Y2H-seq approach (Figure 6). Here, we want to discuss the advantage of the new developed Y2H-seq approach in comparison to the state of the art Y2H-matrix approach in our lab. In the Y2H-matrix approach baits containing the protein of interest were separately screened against individual preys in the array of 14,268 proteins. This was repeated for four replicas. Hence, every prey was mated four times with the baits. Therefore, four prey arrays of 14,268 prey clones and 700 ml bait culture were required (Figure 6). This was laborious work and large quantities of growing yeast and material (e.g. the prey array is contained in 38 384-well MTP) were necessary. In the Y2H-seq approach preys were once grown individually in the array format and then pooled. The pooled preys were mixed with the pooled baits. This mix was distributed in a single 384-well MTP and spotted repeatedly on non-selective medium to allow mating in more than 13,800 distinct colonies. Hence, several thousands separate pooled matings were performed and were kept separated during the procedure. Thus, the sampling, i.e. the number of times a protein pair is tested, in the Y2H-seq approach was higher in comparison to the Y2H-matrix screen where only four matings per prey were performed. To select for interactions the yeast was transferred to selective medium at high density. Diploid positives were selected based on reporter gene activation and the resulting ability to grow on selective media. Hence, growth indicated interaction between at least one bait and one prey of the pool.

In theory every spot contained all interacting baits and preys but not all spots grew on selective medium. This indicated that an interaction between a bait and a prey could not be detected every time it was tested. Therefore, it was important to perform a large number of matings and to allow interactions to take place in a large number of spots. On the other hand this result also suggests that it is not a strong single interaction that will compete with all other interactions in the pool but rather that growing colonies on selective agar will be due to different interactions. However, we spotted individual matings typically more than 120,000 times and obtained 5 to 10 times more yeast colonies on the selective agar of the Y2H-seq screen in comparison to the same number of selective agar in the Y2H-matrix approach (compare Figure 8 and Figure 9). This indicates that interactions are tested more exhaustively in the Y2H-seq approach.

In a classical library approach the preys are pooled similar to the Y2H-seq approach. The crucial factor is that a library prey pool contains random cDNA fragments or open reading frames (ORFs) (Chien et al., 1991; Reboul et al., 2003). Thus, some cDNAs are overrepresented and others underrepresented in the prey library by orders of magnitudes. This limits the sensitivity of the Y2H library approach (Reboul et al., 2003). In contrast, in the Y2H-seq and in the Y2H-matrix

screen an ordered array of individual subcloned and characterized prey clones was used (Worseck et al., 2012). Hence, all proteins are represented well in the prey array and normalized mixtures of preys have been demonstrated to result in up to 100 times higher sensitivity (Reboul et al., 2003).

The comparison of the two screening approaches revealed that 30 % of the primary hits identified in the Y2H-matrix screen were also identified in the Y2H-seq screen (Figure 12). A similar overlap is expected if the Y2H-matrix screen would be simply repeated (Venkatesan et al., 2009). Hence, the primary hits identified in the Y2H-seq and the Y2H-matrix approaches were comparable. Also the number of interactions per bait were comparable between the Y2H-matrix and the Y2H-seq approaches suggesting that neither hubs nor proteins with fewer interactions work better in one of the approaches. In the Y2H-seq approach the prey proteins identified were ranked according to the quantity of reads. Preys with many reads were retested with a high retest success rate (describes the percentage of primary hits positively retested), hence the quantity of reads correlates well with the retest success rate (Figure 15). We calculated the success rates for the top 100, top 200 and so forth ranked preys. Therefore, we were able to compare the number of interacting preys at similar retest success rates between the Y2H-seq and the Y2H-matrix screen. The comparison demonstrated that more preys are identified to be retested in the Y2H-seq screen than in the Y2H-matrix screen at a similar retest success rate (Figure 16). In agreement with the much larger number of growing spots obtained in the Y2H-seq approach, this result indicates that the Y2H-seq method as such has approximately two to four fold higher sampling sensitivity.

In summary, the Y2H-seq approach has an increased sensitivity and thus a higher coverage of PPI data in comparison to the Y2H-matrix approach. In addition, the workload is reduced substantially. Notably, the method has the same specificity as the retest performed is identical in both approaches. The second generation sequencing is cost intensive but the cost will decrease in the future. Hence, the Y2H-seq approach would be an efficient, cost-effective strategy and could complement current efforts in interactome mapping.

5.1.2 Y2H methods utilizing second generation sequencing

Recently the group around Marc Vidal presented a Y2H second generation sequencing approach which largely differs from ours. The key step in the approach is the so called stitch-PCR. In the stitch-PCR the prey and bait sequences encoding the interacting proteins are connected with a linker sequence on one PCR amplicon obtained from the yeast colony (Yu et al., 2011). The stitched PCR product had a 82 bp long linker between the sequences encoding the interacting proteins. The purified stitched PCR products (each for every interacting pair) were pooled and sequenced together on a 454 FLX sequencer. In the presented experiment about 400,000 reads were obtained. Of the sequences obtained 39,000 included the linker and 19,000 additionally parts of the protein sequences required to determine their identity. Hence, 95 % of the sequences obtained were not used because the linker or the protein sequences were missing. The 19,000 reads matched to 2,000 primary hits. Hence, on average 9.5 sequences per prey protein were identified. In reality,

reads are not normally distributed. Some pairs were obtained by many reads and the majority of sequences were obtained by a few if not a single read. Hence, primary hits could not be ranked by the read quantity. Thus, all 2,000 primary hits needed to be retested to result in high-quality interaction data. In the retest 1,318 of the primary hits were validated.

In comparison to our Y2H-seq approach the advantage of the Y2H-stitch PCR approach is that interacting proteins were identified directly from the sequenced reads. The disadvantage is that it exploits only a small amount of the sequence information and it thus not quantitative. Primary hits can not be ranked and therefore all primary hits needed to be retested. Our Y2H-seq approach in contrast allowed ranking of primary hits. About 55 % of the total mappings (~80 million) mapped to NCBI Reference sequence (RefSeq) annotated genes. Even though about 7000 RefSeq annotated genes are mapped at least once, more than 99 % of the mappings were actually used to rank the top 300 hits. The top ranking genes, above a certain sequencing score cut off, are essentially all true interactors. Failure of high ranking genes in the retest can be attributed to wrong clones that were picked from the prey array (Corwin und Woodsmith, personal communication). The ranking allows application of the Y2H-seq approach in other interactome-mapping methods, in particular other binary protein-protein interaction assays, like yeast one hybrid or genetic screens in which pairs of DNA molecules are selected and identified.

The Y2H-seq approach can be developed further if barcode tags are used for a multiplexing sequencing approach. There is a Multiplexing Sample Preparation Oligonucleotide kit available from Illumina were up to 96 samples can be tagged and sequenced on a single flow cell. Baits would be separately screened against the prey array. The sample tested with specific bait would be tagged with a unique oligonucleotide in the PCR reaction prior to sequencing, thus labeling the interactors from each bait. All samples would be sequenced together and the prey sequences would encode a barcode assigning for the interacting bait. This would allow the identification of interactions between bait and prey directly from the sequenced reads, as in the Y2H-stitch approach. However the Y2H-seq approach would still give quantitative information on how reliable the prey interacts with the bait.

Here we discussed different and largely complementary features of our Y2H-seq approach with the recently proposed Y2H stitch-PCR by the Vidal lab, CCSB, Boston (Yu et al., 2011). The Y2H-seq approach can be advanced by using barcode tags for sequencing. This would in principle also allow the identification of interacting bait and prey pairs directly from the sequenced reads at a high throughput. The advantage of the Y2H-seq approach in comparison to the Y2H-stitch seq is the capability to generate a ranked primary hit list that correlates with the probability that two proteins interacts.

5.2 First proteome-wide interaction dataset of PMT and PDeM

Arginine and lysine methylation on proteins are important posttranslational modifications that regulate various cellular processes (Bedford and Clarke, 2009). Progress to date has uncovered

only a small portion of the roles of protein methylation in regulation of protein function in physiological processes (Lee and Stallcup, 2009). Identification of methylated proteins is difficult because the detection methods are limited by the lack of affinity reagents or high quality antibodies (Komyod et al., 2005; Pahlich et al., 2006) and because of limitations in radioactive labeling (Levy et al., 2011). Proteins are methylated by protein methyltransferases (PMTs) and a physical interaction is a requisite for PMT enzymes to methylate target proteins. Thus, substrates ought to be detected by studying direct interactions (Passos et al., 2006a). The Y2H system detects direct interaction and is known to be particular capable to detect transient interactions (Yu et al., 2008). Transient PPI are efficiently detected if the sampling is increased (Vinayagam et al., 2010). Hence, the high sampling sensitivity of the Y2H-seq approach is of particular importance when detecting transient PMT-substrate interactions.

Passos et al. show that PRMT1 substrates can be discovered as interacting proteins in a Y2H approach (Passos et al., 2006a). We extend the set of protein arginine methyltransferases (PRMT) by using protein lysine methyltransferases (PKMT) to detect interacting proteins as potential methylation substrates. Protein demethylases (PDeM) were also tested under the assumption that methylated proteins are demethylated and thus physically interact with PDeM. To detect potential methylation substrates we screened 34 enzymes including PMTs and PDeMs, against a prey array containing representative ORFs of about 57 % of the human protein coding genes. Our study resulted in 523 PPIs between 324 preys and 22 PMTs/PDeMs, representing the first large interaction dataset for enzymes involved in methylation. It will serve as resource to identify and analyze novel methylated proteins.

5.2.1 Known methylated proteins detected in the PMT/PDeM network

To confirm that methylation substrates can be identified by studying PMT- and PDeM-protein interactions we searched for already known methylated proteins in our interaction dataset. As anticipated, we identified the methylation substrates histone H2 and histone H3. Both are known to be lysine methylated (Chang et al., 2007; Jenuwein and Allis, 2001). In addition, histone 3 is methylated on arginine 3, 9 and 18 (Kouzarides, 2007) (Table 4). Beside histones, we found also non-histone arginine methylation substrates like RNA binding proteins including SERBP1 (Passos et al., 2006a), EWSR1 (Belyanskaya et al., 2001; Kim et al., 2008; Pahlich et al., 2008) and QKI (Cote et al., 2003). We identified SERBP1 interacting with PRMT8 although SERBP1 is known to be methylated by PRMT1 (Passos et al., 2006a). Except for SERBP1 all PRMT8 interaction partners interact with PRMT1, too (Figure 24). This is probably due to the 80 % sequence similarity of PRMT8 to PRMT1 (Kim et al., 2008). Hence, we conclude that SERBP1 interacts with PRMT8 and PRMT1, but we missed the later in our screen (false negative result). Furthermore, we suggest that SERBP1 might be methylated by both PRMT1 and PRMT8, like EWSR1 (Kim et al., 2008). The isoform one of QKI (QKI-5) is known to be methylated but the enzyme responsible is elusive (Cote et al., 2003). We found QKI interacting with CARM1 and

suggest methylation by CARM1 (see discussion below). We also identified arginine methylated heterogeneous nuclear ribonucleoproteins (hnRNP) including SYNCRIP (Passos et al., 2006b), FUS (Rappsilber et al., 2003; Tradewell et al., 2012) and HNRNPH2 (Lee and Stallcup, 2009; Najbauer et al., 1993). FUS was identified as SUV39H1 interaction partner what is not in agreement with the known arginine methylation site on FUS. However, FUS might be methylated on a lysine residue, too. As it is methylated by PRMT1 and we did not identify FUS interacting with PRMT1 this is a false negative result. HNRNPH2 is proposed to be methylated on arginine 217 (Swissprot). We found HNRNPH2 interacting with PRMT6, suggesting that HNRNPH2 is methylated by PRMT6. Furthermore, we identified PRMT8 and KRT7 both are known to be arginine methylated (Kim et al., 2008; Ong et al., 2004; Sayegh et al., 2007).

As the only non-histone lysine methylated protein we found WIZ interacting with SUV39H1/2 and AOF2. WIZ contains several zinc fingers and was identified as a target for methylation by EHMT2 (Rathert et al., 2008). EHMT2, SUV39H1 and SUV39H2 are lysine methyltransferases and are known to regulate the methylation status of mono-, di- or trimethylated histone H3 (Bannister et al., 2002; Jacobs and Khorasanizadeh, 2002; Rathert et al., 2008). Interestingly the lysine 305 methylated on the WIZ protein functions as a methyl specific binding site for heterochromatin protein 1 (HP1), in analogous manner to the histone H3 lysine 9 trimethylation binding site for HP1 (Rathert et al., 2008) (see discussion below).

Progress to date has only discovered a small number of methylated proteins. In the swissprot database are 69 arginine sites and 25 lysine sites listed on non-histone proteins. Taken into account the small number of known methylated proteins, the incompleteness of the prey array and that not all protein contained in the network are methylation substrates, the chance to identify known methylated proteins was relatively low. Still, we were able to detected eight non-histone proteins known to be methylated on arginine residues. Furthermore, we were able to detect one non-histone lysine methylated protein. This suggests that the Y2H screen performed efficiently identified potential arginine and lysine methylation substrates.

5.2.2 New methylation substrates characterized in the methylation assay

To identify novel methylation substrates from the set of interacting proteins we developed a methylation assay. Methylation is mostly discovered by radioactive labeling or mass spectrometry (Levy et al., 2011). Here, we established an *in vitro* assay using radioactive labeled methyl donor ^{3}H *S*-adenosylmethionine (^{3}H SAM, ^{3}H AdoMet) (Lee et al., 2004) and detected the incorporation of radioactive material by scintillation counting. SYNCRIP and PABP1 (control) were detected to incorporate radioactive material indicating methylation (Figure 25B). To determine novel sites of methylation we performed the same assay with non radio labeled AdoMet and identified methylation sites by LC-MS/MS.

PRMT1 is the major active arginine methyltransferase in cultured cells (Bedford and Clarke, 2009; Pawlak et al., 2002; Tang et al., 2000b) and hence we focused on PRMT1 interacting

proteins. We detected methylation on potential PRMT1 substrates including SYNCRIP, OFCC1, WDR42A and SAMD3 but failed to monitor arginine methylation of SERBP1. SERPB1 has two conserved RG-rich regions that are targets for methylation *in vitro* and bind to PRMT1 (Passos et al., 2006a), but corresponding peptides were not identified in our approach, neither methylated nor unmethylated. This is probably due to the fact that before MS analysis the probes were trypsin digested. Trypsin cleaves peptide chains mainly at the carboxyl side of arginine and lysine. Hence, RG-rich regions were cleaved into very small peptides and therefore may have escaped detection in the MS. To get larger peptides the digesting enzyme needs to be changed or SERBP1 has to be partially digested. SAMD3 interacted with PRMT1 and AOF2. Hence, we expected arginine and lysine methylation. We incubated SAMD3 with HEK cell lysate overexpressing PRMT1 but SAMD3 can be lysine methylated by one of the several endogenous PKMTs in the HEK cell lysate. SAMD3 was detected with sequence coverage of 27 % and to be methylated on lysine 157. OFCC1 is monomethylated on R105, R113 and R195. The function of OFCC1 is largely unknown. A report in humans identified incomplete homologous transcription of the gene within a locus linked to orofacial clefting hereditary disease (Davies et al., 2004). We were able to detected 15 mono- and dimethylated arginine sites on SYNCRIP. Five methylation sites were located in the RRM motifs (162 to 408) and nine in the RG-rich (448 to 559) C-terminal region of SYNCRIP. N-terminal of the RG-rich region we identified methylated arginine 409. Our methylation data are supported by a previous study by Passos et al. showing that the C-terminal region (389 to 623) of SYNCRIP is methylated. *In vivo* experiments inhibiting the methylation with the global methylation inhibitor Adox causes a change in SYNCRIP localization from strictly nuclear to partially cytoplasmic. Hence, SYNCRIP methylation in its C-terminal region is important for its nuclear localization (Passos et al., 2006b).

We also detected methylation on CARM1 interacting proteins QKI and DNAJA3. DNAJA3 was classified as a molecular cochaperone stimulating the ATPase activity of heat shock protein 70 chaperon (Hsp70). It plays a critical role in protein folding, degradation and multimeric complex assembly (Garrido et al., 2003). DNAJA3 is predominantly localized to the mitochondria (Kurzik-Dumke et al., 1998; Syken et al., 1999). However, the reported DNAJA3-protein interactions and functions are primarily non mitochondrial. It interacts directly with cytosolic and nuclear proteins and participates in signaling pathways mediated by Jak2 (Sarkar et al., 2001), NFκB (Cheng et al., 2002), (Trentin et al., 2001) and Trk receptor tyrosine kinases (Liu et al., 2005) modulating e.g. cell death and proliferation. Methylation of DNAJA3 could thus be involved in many cellular functions, like localization, mediating PPI, functional activating of DNAJA3 or interacting proteins, degradation or protein folding. To clarify the function of DNAJA3 methylation further experiments have to be performed. SPIN2B was also found to be methylated on arginine 15, 53, 100 and 200. SPIN2B was not identified in the Y2H analysis, but SPIN1 was detected interacting with PRMT6. SPIN2B was used in the methylation assay as it was successfully isolated from *E. coli* in sufficient amounts. SPIN1 and SPIN2B have a sequence similarity of 73 % and arginine 15, 53 and 200 are

conserved in SPIN1 whereas arginine 100 is converted to a lysine in SPIN1. SPIN2B was functional cloned by Flechter et al. and identified as a nuclear anti-apoptotic protein with roles in cell cycle progression (Fletcher et al., 2002). Biochemical functions and mechanism of the spindling protein family are largely unknown. More research needs to be done to clarify the function of the spindlin proteins and the effect of methylation on spindlin proteins. QKI isoform one (QKI-5) is methylated on two arginine residues and discussed in a separate section. The WDR42A protein was methylated on seven arginine residues and we will discuss these findings below.

In summary, we detected methylation of SPIN2B, OFCC1, WDR42A, QKI, DNAJA3, SAMD3 and SYNCRIP. We conclude that more unknown methylation substrates are contained in the dataset. With the identification of novel arginine and lysine methylation sites on proteins identified in the Y2H PMT/PDeM interaction network we exemplarily validated our approach.

5.3 Proteins interacting with PDeMs

5.3.1 AOF2 interacts with a variety of potential non-histone demethylation substrates

AOF2 (LSD1, KDM1A) is a well studied histone lysine demethylase (Shi et al., 2004). We identified AOF2 interacting with 182 nuclear and cytoplasmic proteins. A significant fraction of 22 proteins are cytoskeleton proteins and seven of those are keratins, indicating a function of AOF2 in structural constitution of the cytoskeleton (Figure 21 and Figure 22). Hence, AOF2 might have cytoplasmatic in addition to its nuclear functions. This is in agreement with the observation that AOF2 is known to demethylate non-histone substrates as p53, MYPT1 and DNMT1 (Cho et al., 2011; Esteve et al., 2011; Huang et al., 2007). We identified WIZ that is known to be methylated (Rathert et al., 2008) interacting with AOF2. As candidate demethylation substrates AOF2 interacting proteins present a large set of potentially methylated proteins.

5.3.2 JMJD6 interacts with spliceosomal proteins

At the beginning of this study JMJD6 was identified as the first arginine demethylase (Chang et al., 2007). Therefore, we selected JMJD6 and studied JMJD6-protein interactions. Later it was shown that JMJD6 catalyses lysyl-hydroxylation of U2AF65 (Webby et al., 2009) and thus likely not exhibits demethylation. Hence, the enzymatic activity of JMJD6 is under debate as its cellular function.

Webby at al assayed for proteins interacting with JMJD6 using affinity purification coupled to mass spectrometry (AP-MS) and identified 39 proteins. A substantial proportion (22 of 39) of the potential JMJD6-interacting proteins is connected to RNA metabolism, processing and splicing, with 25 % being nuclear proteins possessing arginine-serine domains (12 of 39). They showed that U2AF65 is hydroxylated by JMJD6 (Webby et al., 2009). U2AF65 is required for mRNA splicing (Sickmier et al., 2006) and modulation of splice-site recognition by U2AF65 influences alternative

splicing (Hastings et al., 2007). In JMJD6 knock down HeLa cells the alternative splicing pattern of the endogenous tumor antigen MGE6 and the α-tropomysin gene is altered through an increase in the amount of exon skipping (Webb et al., 2008). This study suggests that hydroxylation affects intron 3´splice site strength of U2AF65 and possibly other splicing factors and argues strongly against a role of JMJD6 in protein demethylation (Webb et al., 2008). We identified 26 proteins directly interacting with JMJD6 in the Y2H screens. Noticeable, four of the 26 proteins interacting with JMJD6 contained RG-rich regions including RSRC1, U2AF1, ANKRD5 and BRD4 (Figure 20) whereof RSRC1 and U2AF1 have a SR-rich region. RG-rich regions are often found in RNA binding proteins (Bedford and Richard, 2005; Liu and Dreyfuss, 1995). We identified RNA binding proteins NHP2 (Wang and Meier, 2004) and LARP7 (He et al., 2008) interacting with JMJD6. Furthermore, we identified five spliceosomal proteins interacting with JMJD6 including SLU7 (Chua and Reed, 1999), RSRC1 (Cazalla et al., 2005), CCNL1 (Dickinson et al., 2002), PRPF38A (Blanton et al., 1992) and U2AF1 (U2AF35) (Zuo and Maniatis, 1996). Notably, in agreement with our finding and the JMJD6 complexes detected by Webby et al., the U2AF35-U2AF65 interaction is well characterized (Hegele et al., 2012; Rudner et al., 1998). The direct interaction of JMJD6 with five spliceosomal factors and two RNA binding proteins suggest a direct involvement of JMJD6 in splicing. JMJD6 could modulate the function of splicing factors to contribute in the regulation of alternative splicing (Webby et al., 2009).

5.4 Proteins interacting with protein lysine methyltransferases

We identified a representative set of interaction for the lysine methyltransferases SUV39H1, SUV39H2, WHSC1L1 and SMYD1. Those enzymes are known to exclusively methylate histones (Kim et al., 2006a; O'Carroll et al., 2000; Robin et al., 2007; Tan et al., 2006). Whereas SETD7, SMYD3, SETD6 and EHMT2 are known also to methylate non-histone proteins (Huang and Berger, 2008; Kunizaki et al., 2007; Levy et al., 2011; Rathert et al., 2008). We discovered many nuclear and cytoplasmic proteins interacting with PKMTs. This supports the assumption that other PKMTs methylate non-histone substrates, too. Anyway, PKMTs seem to be involved in diverse functions and not exclusively methylate histones. Especially the cytoplasmic proteins are interesting and indicate a completely unexploited field of PKMT functions.

5.4.1 SUV39H1 preferentially interacts with zinc finger (ZnF) containing proteins and methylates the ZnF protein WIZ

We identified the WIZ (widely-interspaced zinc finger-containing protein) protein (Figure 28) interacting with lysine methyltransferase SUV39H1/2 and lysine demethylase AOF2. WIZ was discovered as EHMT2 (G9a) substrate based on sequence similarity to a known substrate (Rathert et al., 2008). In detail, EHMT2 generates mono- and dimethylation on histone H3 lysine 9 (H3K9) (Collins et al., 2005). On this basis an optimal target sequence for EHMT2 was developed and used

in a proteome-wide search. Thereby, non-histone targets of EHMT2 including WIZ were identified (Rathert et al., 2008). As H3K9 is mono- and dimethylated by EHMT2 (Peters et al., 2003) we suggest that WIZ is mono- and dimethylates by EHMT2, too. H3K9 is trimethylated by SUV39H1 (Peters et al., 2003). Therefore and because we identified SUV39H1 interacting with WIZ we suggest that WIZ is trimethylated by SUV39H1. Additional, we identified AOF2 interacting with WIZ (Figure 23). As AOF2 is able to remove mono- and dimethyllysine (Shi et al., 2004) we propose that AOF2 removes mono- and dimethylation of WIZ (Figure 28B). JMJD1A and JMJD2C demethylate H3K9me1/2 and H3K9me3, respectively (Loh et al., 2007; Ueda et al., 2006). Hence, trimethylation on WIZ protein could be removed by JMJD2C. Trimethylation of WIZ and H3K9 recruits the heterochromatin protein 1 (HP1). HP1 binds through its chromo domain to methylated but not unmethylated WIZ and H3K9 (Figure 28B) (Bannister et al., 2001; Rathert et al., 2008). WIZ is localized at the chromatin in a complex with EHMT2, EHMT1, SETDB1 and SUV39H1 methylating H3K9 (Fritsch et al., 2010; Ueda et al., 2006). Hence, we suggest that WIZ is involved in the recruitment of HP1 to chromatin and functions in chromatin regulation by mediating gene silencing (Bannister et al., 2001). The methylation specific binding of HP1 to non-histone targets shows that binding can be regulated by methylation on non-histone proteins. Furthermore, methylation of non-histone proteins can be recognized by methyl binding domains, similar to what is known for histone methylation. Methylation of non-histone chromatin associated proteins adds a new layer to epigenetic signaling that could be important for regulating the activity and interaction of chromatin factors (Rathert et al., 2008). Here we identified enzymes likely responsible for the reversible methylation of the zinc finger protein WIZ.

As mentioned above, WIZ contains 11 zinc finger domains (Figure 28A). Proteins interacting with SUV39H1 are enriched for zinc finger C2H2-type domains, Krueppel-associated box (KRAB) and SCAN domain (Figure 22). Co-enrichment of those domains is expected because KRAB and SCAN domain are associated with C2H2-type zinc finger (Peng et al., 2000; Williams et al., 1995). C2H2-type zinc finger proteins are present in many transcriptional factors and DNA-binding proteins (Klug and Schwabe, 1995; Laity et al., 2001; Leon and Roth, 2000). The KRAB as well as the SCAN function as DNA-binding dependent transcriptional repression modules (Peng et al., 2000; Williams et al., 1995). This is in agreement with the enrichment in transcriptional regulator activity and DNA-binding function of proteins interacting with SUV39H1 (Figure 21). We conclude that many transcription factors (TF) interact with SUV39H1. In general terms, the interaction of transcription factors with SUV39H1 could imply, for example, that TFs are methylated or that SUV39H1 acts on the chromatin and recruits TFs to the chromatin. If TFs are methylated their activity may be modulated by proteins containing methyl binding domains. Hence, methylated TFs would recruit methyl binding domain containing proteins as co-regulators to the DNA to modulate transcription (Figure 29A) (Taverna et al., 2007). Supported by the fact that TFs are known to be arginine methylated. For example, CARM1 contributes to transcriptional regulation by methylation of E1A binding protein p300 (EP300, p300) (An et al., 2004), CREB

binding protein (CREBBP, CBP) (Chevillard-Briet et al., 2002; Xu et al., 2001) and nuclear receptor coactivator 3 (NCOA3, SRC-3, AIB1) (Bedford and Richard, 2005; Stallcup, 2001). This suggests a new layer of lysine methylation dependent signaling that is important for regulating transcriptional activity of chromatin.

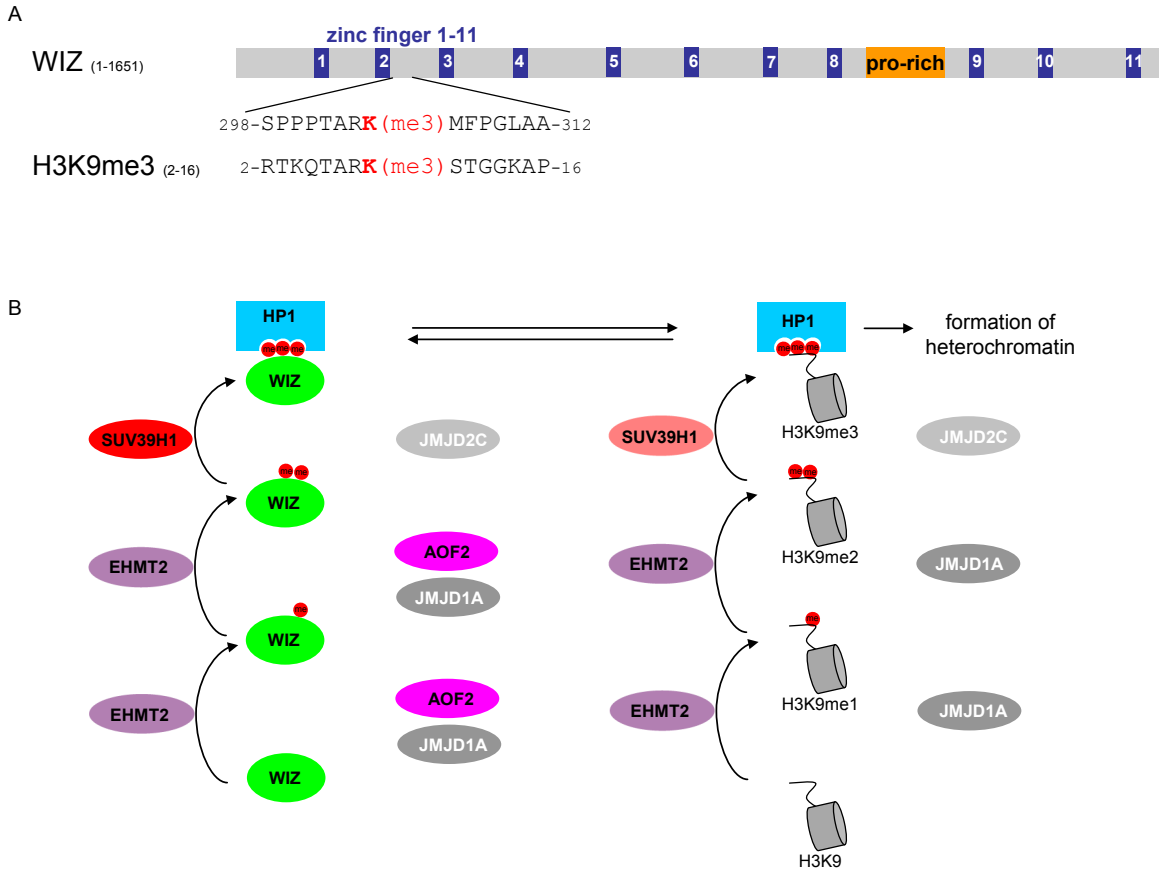

Figure 28: WIZ methylation and demethylation

We identified interactions between AOF2 and SUV39H1 with WIZ. (A) Schematic representation of WIZ protein, containing 11 zinc fingers and a proline-rich (pro-rich) region. The methylated sequence identified on WIZ by Rathert et al. is indicated and compared to the methylated sequence of the histone H3 at lysine 9 (H3K9). (B) Model of WIZ methylation and demethylation in analogy to H3K9 methylation and demethylation. WIZ is mono- and dimethylated by EHMT2 (Rathert et al., 2008) and we suggest that it is trimethylated by SUV39H1. Upon methylation HP1 binds to the WIZ protein (Rathert et al., 2008). WIZ is demethylated by JMJD2C and JMJD1A/AOF2. In analogy, H3K9 tail is mono- and dimethylation by EHMT2 and trimethylation by SUV39H1 (Rea et al., 2000). Upon trimethylation HP1 binds to the H3K9me3, a hallmark of heterochromatin (Bannister et al., 2001). H3K9me3 is demethylated by JMJD2C and JMJD1A (Loh et al., 2007).

Alternatively, SUV39H1 could bind TFs and recruit them to the chromatin. There SUV39H1 methylates histone tails regulating the chromatin state and thus couple TF activation or repression to the histone modification state. In other words, the chromatin state and the transcription would be co-regulated through recruitment of TFs by SUV39H1 (Figure 29B). In agreement with such a proposal, CARM1 functions as coactivator for several DNA-binding transcription factors including nuclear receptors, p53 (An et al., 2004), nuclear factor kappa B (NFκB) (Covic et al., 2005; Hassa et al., 2008), ß-catenin/LEF (Koh et al., 2002), E2F transcription factor 1 (E2F1) (Frietze et al.,

2008) and cyclin E1 (CCNE1) (El Messaoudi et al., 2006). Furthermore, nuclear receptor transcriptional activity, e.g. of human androgen receptor (AR) and human estrogen receptor (ER)α, is coupled to the chromatin by CARM1 (arginine methyltransferase) (Lee et al., 2006; Lee et al., 2005d; Lee and Stallcup, 2009). AOF2 interacting proteins were enriched in similar GO categories as SUV39H1. Therefore, we suggest a similar function but counteracting SUV39H1 (Figure 29A und B). In summary, we find a significant number of TFs interacting with the chromatin modifying enzymes SUV39H1 and AOF2.

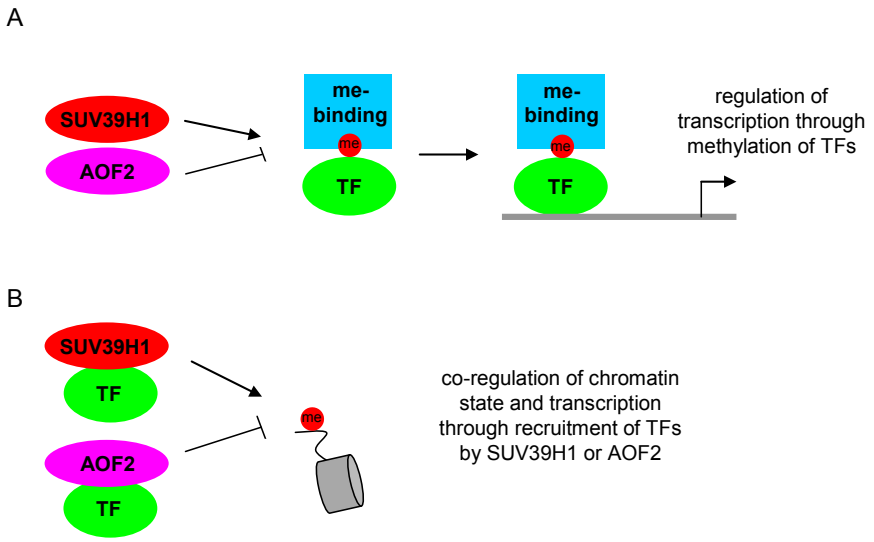

Figure 29: Two alternative models for transcription factor (TF) binding to SUV39H1 and AOF2

Schematic representation of how TF binding to SUV39H1 and AOF2 could be involved in regulation of transcription and chromatin state. (A) TFs are methylated and demethylated by SUV39H1 and AOF2, respectively. Methylated TF are bound by methyl binding domains and recruited to the DNA where they function in regulation of transcription. (B) TFs bind SUV39H1 and AOF2 and co-recruit them to the chromatin. Hence, histones are methylated or demethylated and heterochromatin or euchromatin is formed.

5.5 Proteins interacting with protein arginine methyltransferases

We discovered 164 proteins interacting with PRMTs. A representative set of interaction were identified for PRMT1/8, CARM4, PRMT5 and PRMT6. Especially for PRMT6 we were able to detect many new interacting proteins and expand the current knowledge significantly (Bedford and Clarke, 2009; Lee and Stallcup, 2009).

5.5.1 CARM1 is the putative methyltransferase for the STAR protein quaking

Quaking (QKI) belongs to the family of signal transduction and activation of RNA (STAR) proteins (Ebersole et al., 1996). These RNA binding proteins share a single extended K homology (KH) domain (Vernet and Artzt, 1997). Several STAR family proteins, SAM68, SLM-1, SLM-2, GRP33 and QKI-5 are known to be methylated. SAM68, SLM-2 and GRP33 are methylated by PRMT1 whereas the methyltransferase methylating SLM-1 and QKI-5 is unknown (Cote et al.,

2003). We identified QKI interacting with CARM1 in the Y2H screens (Figure 23). Furthermore, in the methylation assay coupled to LC-MS/MS we identified the R43 monomethylated and the R242 mono- and dimethylated (Figure 30A).

There are three alternative quaking isoforms QKI-5, QKI-6 and QKI-7 (Kondo et al., 1999; Li et al., 2002). They share the same N-terminus but have a distinct C-terminus determining their subcellular localization. QKI-6 and QKI-7 are localized in the cytoplasm whereas a nuclear localization signal (NLS) in the C-terminal of QKI-5 allows its import into the nucleus (Hardy et al., 1996) (Figure 30A). Thus, QKI-5 is found predominantly in the nucleus. However, QKI-5 can shuttle between the nucleus and the cytoplasm (Wu et al., 1999).

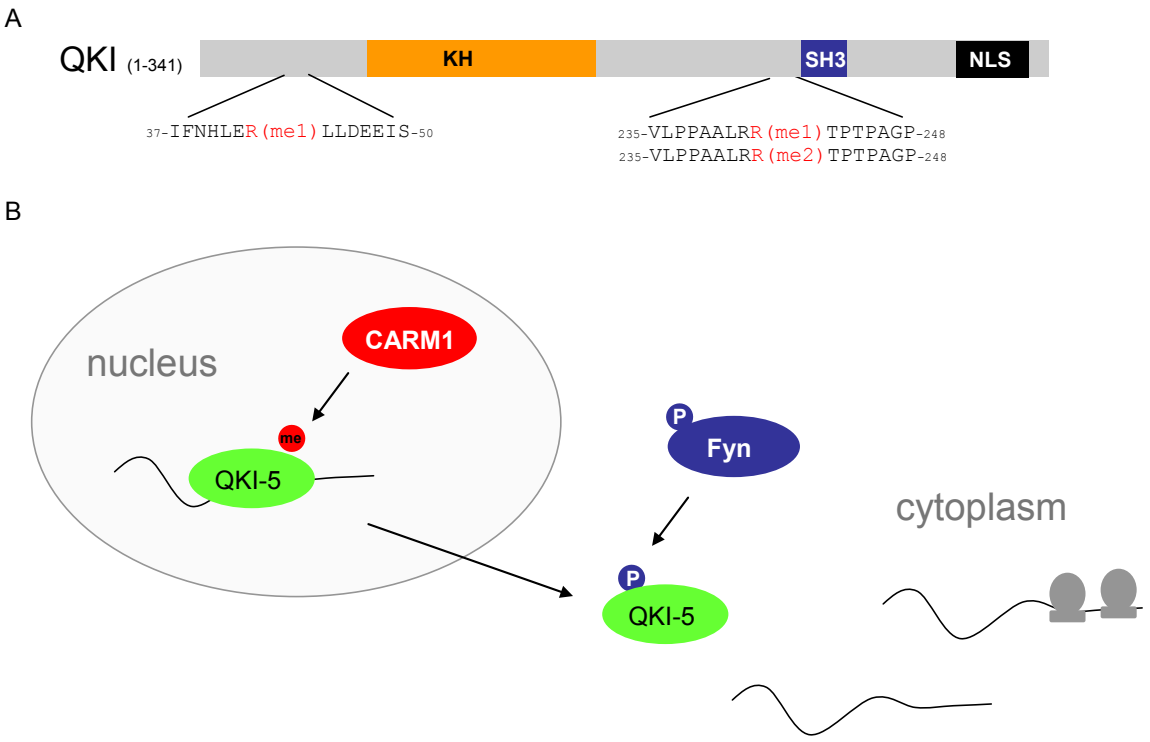

Figure 30: Methylation dependent shuttling between the nucleus and cytoplasm of QKI in analogy to Sam68

(A) Schematic representation of QKI-5. QKI contains the elongated K homology (KH) domain, a Src homology 3 (SH3) and nuclear localization signal (NLS). The monomethylated arginine 43 and the mono- and dimethylated arginine 242 found in the LC-MS/MS approach are indicated. (B) Model of QKI-5 methylation and phosphorylation in analogy to Sam68 methylation and demethylation. Methylated QKI-5 is localized in the nucleus and binds RNA, if QKI is demethylated it partially relocalized into the cytoplasm. There QKI-5 is phosphorylated and releases the target RNA (Zhang and Cheng, 2003).

Likewise Sam68, a well studied STAR family member, QKI is predominantly localized in the nucleus. Upon PRMT1 knockout or inhibition a significant fraction of Sam68 is localized to the cytoplasm. Nuclear localization of Sam68 requires the NLS and methylation of the RG-repeats 100 amino acids N-terminal of the NLS (Cote et al., 2003). Methylated Sam68 is localized in the nucleus and demethylation promotes cytoplasmatic localization. QKI and Sam68 have a similar NLS (Wu et al., 1999) and about 100 amino acids N-terminal of the NLS they are both methylated.

Therefore, we suggest a similar shuttling mechanism for QKI-5. QKI-5 occurs methylated and predominantly nuclear in the cell. Hence, methylated QKI-5 is probably localized in the nucleus and upon demethylation it is localized to the cytoplasm, similar to Sam68.

An important function of shuttling STAR proteins is transporting RNA, and in this sense, the methylation could function as regulator of RNA transport (Wu et al., 1999). QKI-5 binds e.g. the mRNA of myelin basic protein (Zhang et al., 2003). Upon demethylation QKI would escorts MBP mRNA out of the nucleus. In the cytoplasm QKI-5 is phosphorylated by the tyrosine kinases Src or Fyn. Phosphorylation decreases the affinity for MBP mRNA (Zhang et al., 2003). Hence, the MBP mRNA is released in the cytoplasm. After mRNA release QKI-5 could then be recycled back to the nucleus upon methylation (Figure 30B).

In summary, we could recapitulate the methylation of QKI found by Côté et al.. Additionally, we identified arginine 43 monomethylated and arginine 242 mono- and dimethylated, that are most likely modified through interaction with CARM1. The QKI RNA binding is regulated by phosphorylation (Zhang et al., 2003) and we suggest that the RNA export is regulated by QKI-5 methylation.

5.5.2 Methylation of the SOCS box protein SPSB2 by PRMT1 maybe involved in regulating ubiquitination and degradation of target proteins

SPSB2 belongs to the SPRY domain–containing SOCS box protein family consisting of SPSB1, SPSB2, SPSB3 and SPSB4. They are composed of a central SPRY protein interaction domain and a C-terminal SOCS box (Hilton et al., 1998; Nicholson and Hilton, 1998). We identified SPSB1 and SPSB2 interacting with PRMT1 (Figure 24). SPSB2 contains a RG-rich region (Figure 20) in the SPRY domain which is partially conserved in SPSB1 (Figure 31A). Other proteins identified interacting with PRMT1 also contained RG-rich regions including FUS, SERBP1, EWSR1 and SYNCRIP (Figure 20). Those proteins are known to be methylated by PRMT1 (Belyanskaya et al., 2001; Passos et al., 2006a; Passos et al., 2006b; Rappsilber et al., 2003). SPSB2 is the only protein interacting with PRMT1 and containing a RG-rich region but not known to be methylated. Therefore and because PRMT1 is known to methylate RG-rich regions (Bedford and Richard, 2005; Lee and Stallcup, 2009; Liu and Dreyfuss, 1995) we suggest that the conserved arginines in the RG-rich region between SPSB1 and SPBS2 are methylated (Figure 31A). However, our attempts to express and purify SPSB2 for *in vitro* methylation assays failed.

SPSB2, like other SOCS box-containing proteins, functions as substrate receptors in the E3 ubiquitin ligase complex (Kuang et al., 2010). E3 ligases define substrate specificity and covalently attach ubiquitin to lysine side chains of the substrate targeting the proteins for proteasomal degradation (Piessevaux et al., 2008). SOCS box domain mediates the interaction with Elongine C (Aso et al., 1996; Kamura et al., 1998; Kibel et al., 1995; Zhang et al., 2001). Elongine B binds Elongine C and this dimer acts as linker that bridges the substrate recognized by the SOCS box protein to Cullin scaffold protein (Kamura et al., 2001). Cullin in turn recruits a

RING-finger-containing protein Rbx, thereby completing the assembly of the E3 ligase complex (Iwai et al., 1999; Kamura et al., 2001) (Figure 31B). SOCS proteins target various substrates via their additional protein-protein interaction modules to the E3 ligase complex (Kamura et al., 2001; Kamura et al., 1998; Stebbins et al., 1999; Zhang et al., 1999) such as the SPYR domain in the case of SPSB2 (Hilton et al., 1998). The SPRY domain of SPSB1 and SPSB2 interacts with Par-4 (Masters et al., 2006) and iNOS (Kuang et al., 2010). Kuang et al. showed that SPSB2 functions as an adaptor protein in E3 ubiquitin ligase complex and recruits iNOS to the E3 ligase targeting it for proteasomal degradation (Figure 31B) (Kuang et al., 2010).

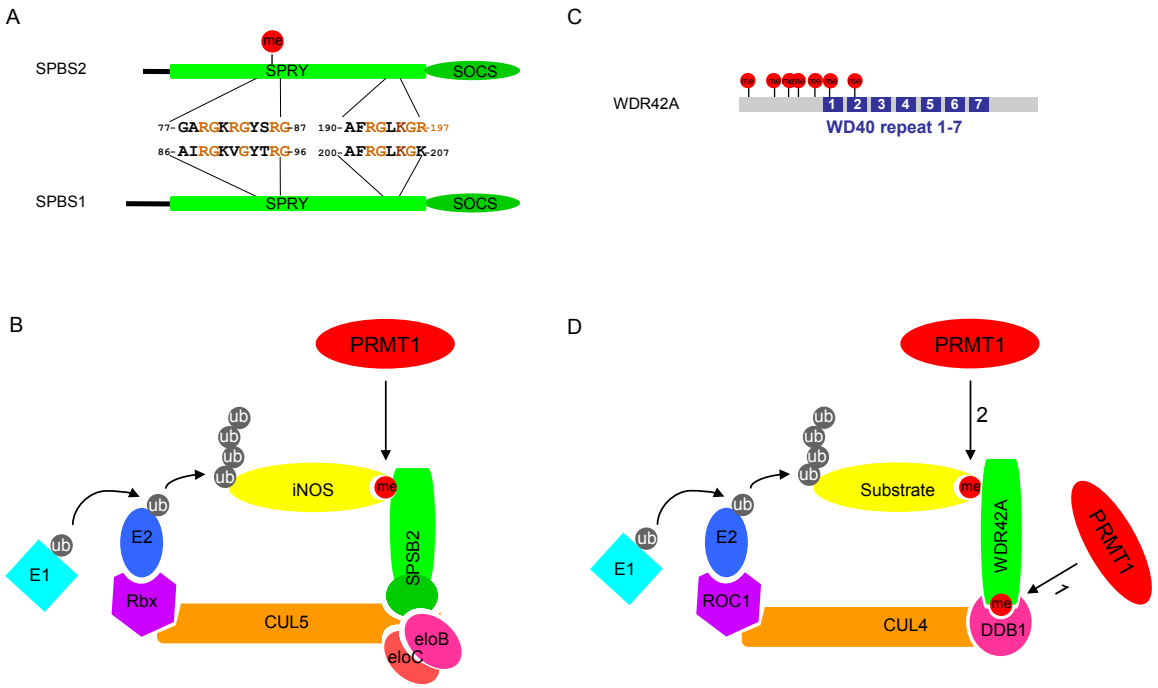

Figure 31: SPSB2 and WDR42A (DCAF8) are adaptor proteins of the E3 ubiquitin ligase complex and recruit substrates for ubiquitination and subsequent degradation

(A) Schematic representation of paralogous SPSB1 and SPSB2. The RG-rich regions in the SPRY domain are indicated as well as potential methylation sites. The C-terminal contains the SOCS domain. (B) In our model, SPSB2 is methylated by PRMT1 in the SPRY domain. SPSB2 interacts with iNOS via SPSB2 SPRY domain and binds to eloign BC complex via its SOCS box, recruiting Cullin 5 and Rbx to form an active E3 ubiquitin ligase complex that polyubiquitinates e.g. iNOS, targeting it for proteasomal degradation (Kuang et al., 2010). The E1 ubiquitin-activating and E2 ubiquitin-conjugating enzyme are indicated. (C) Schematic representation of WDR42A including the seven monomethylated arginine sites found in the LC-MS/MS approach and the seven WD40 repeats. (D) WDR42A is methylated by PRMT1. WDR42A interacts with the Cul4-DDB1-ROC1 ubiquitin ligase that polyubiquitinates substrates targeting them for proteasomal degradation (Angers et al., 2006; He et al., 2006; Lee and Zhou, 2007).

The ubiquitination and proteasomal degradation of iNOS needs to be regulated. The SPRY domain provides one determinant for substrate recognition. Additional posttranslational modifications of the SPRY domain may be required for recruitment and proper representation of substrates to the E2 ubiquitin-conjugating enzyme. As the methylation occurs in the SPRY domain responsible for binding to iNOS we suggest that the binding of iNOS to SPSB2 may be influenced

through methylation. Methylation has the potential to regulate protein-protein interaction, for example Sam68 interactions. Methylation of Sam68 decreases SH3 domain binding, but not WW domain binding, allowing modulation of specific PPIs (Bedford et al., 2000). Another example of methylation dependent protein interactions are those mediated by the methyl recognition motifs such as the tudor domain of SMN which binds methylated SmD3, SmD1 and SmB (Brahms et al., 2001). As one possibility we suggest that interaction of the SPRY domain with the iNOS and other potential ubiquitination substrates are influenced through methylation (Figure 31B).

In summary, we suggest that the RG-rich region of SPSB2 is methylated by PRMT1. Furthermore, SPSB2 functions as substrate receptor of the E3 ligase complex (Kuang et al., 2010). We propose that methylation of SPSB2 modulates the binding of proteins to the E3 ligase complex. Hence, it bears the potential to regulate ubiquitination and degradation of target proteins.

5.5.3 The E3 ligase complex associated WD40 domain containing protein WDR42A is methylated on multiple arginine sites by PRMT1

Interestingly we observed an enrichment of WD40 repeat domain containing proteins interacting with PRMT1 including WD repeat and FYVE domain-containing protein 3 (WDFY3), target of rapamycin complex subunit LST8 (MLST8) and WD repeat-containing protein 42A (WDR42A) (Figure 22). WDR42A binds to DDB1 and therefore is also known as DDB1 and CUL4 associated factor 8 (DCAF8) (Angers et al., 2006). CUL4 is one of the seven cullins assembling to the E3 ligase complex. The substrate recruiting mechanism of CUL4 is poorly understood. CUL4 was identified interacting through its N-terminus with DDB1 (Petroski and Deshaies, 2005). In 2006 four different research labs identified that DDB1 binds to several WD40 repeat containing proteins and function as a linker to recruit the WD40 repeat containing proteins as receptor to the E3 ligase of the ubiquitin system (Angers et al., 2006; He et al., 2006; Higa et al., 2006; Jin et al., 2006). Hence, WD40 repeat containing proteins serve as substrate receptor to recruit proteins for ubiquitination to the E3 ligase (Lee and Zhou, 2007). The structural basis for DDB1 binding to the diverse WD40-containing proteins is mostly unknown (Jin et al., 2006). A special WDXR motif was identified to play a critical role in binding but it is unlikely that the WDXR motif is the only structural determination for DCAFs associated with DDB1 (Angers et al., 2006; He et al., 2006; Jin et al., 2006; Lee and Zhou, 2007). To target distinct cellular substrates for ubiquitination the CUL4-DDB1 core complex needs to maintain efficient and timely assembly with different DCAFs. It is suggested that DDB1 uses different ways to interact with DCAFs and it seems likely that multiple surfaces in DCAF proteins are likewise used to interact with DDB1, probably by also employing residues outside of the tandem WDXR motif (Jin et al., 2006).

We identified seven methylated arginine residues on WDR42A (Table 5 and Figure 31C) which could be a critical factor for WDR42A function such as binding to DDB1. If the side-chain nitrogen atoms of the arginine residues are methylated hydrogen-bonding is disturbed (Gary and Clarke, 1998). This can prevent formation of hydrogen-bonds between proteins and therefore

disturb PPIs, as described for Sam68 (Bedford et al., 2000). Hence, extensive methylation of WDR42A could regulate the binding between WDR42A and DDB1. In addition, methylation could regulate the assembly of DCAFs to different time points and thereby regulate the recruitment of distinct cellular substrates for ubiquitination. Hence, methylation of DCAFs could regulate the activity of CUL4-DDB1-DCAF ubiquitin ligase. This would be an additional structural determination of WDR42A to influence the interaction with DDB1 outside the WDXR motif (arrow 1 in Figure 31D).

Another function of WDR42A methylation could be the recruitment of substrates, similar to SPSB2. While DCAFs provide one determinant for substrate recognition, additionally posttranslational modifications of DCAFs may also be required for recruitment and proper representation of substrates to the E2 enzyme (Lee and Zhou, 2007). Methylated proteins are bound by methyl recognition domains like the tudor domain (Brahms et al., 2001). The methylated DCAFs could specifically recruit methyl binding domain containing substrates. Methylation can also decrease binding affinity, as described for Sam68 (Bedford et al., 2000). Hence, methylation of WDR42A could regulate binding of ubiquitination substrates by decreasing affinity of specific domains and increasing affinity for methyl recognition domains (arrow 2 in Figure 31D).

In summary, methylation of WD40 containing DCAFs can have functions in DDB1-DCAF interactions and regulate the activity of CUL4-DDB1-DCAF ubiquitin ligase complex. We have identified WDR42A/DCAF8 as interaction partner of PRMT1 and mapped seven novel arginine methylation sites. Extensive methylation can alter WDR42A function and its interaction behavior critically affecting DCAF function.

5.6 Summary and further directions

This study presents a novel Y2H approach which identifies interacting proteins utilizing second generation sequencing. The Y2H second generation sequencing (Y2H-seq) approach reduces the workload and increases the sensitivity in comparison to the state of art Y2H-matrix protocols. Especially, transient PPI are efficiently detected in the Y2H-seq approach as the sampling is increased which is of particular importance when detecting transient PMT-substrate interactions. The capability to rank the identified hits because of a quantitative readout allows application of the approach to various interaction assays. This also allows identification of interacting proteins directly from the quantification of obtained reads. Furthermore, the Y2H-seq can be improved in terms of throughput and cost by multiplexing.

In a proteome-wide screen we identified 523 interactions between 324 prey proteins and 22 enzymes associated with methylation. Despite the few proteins known to be methylated and the incompleteness of the prey array we identified 11 prey protein substrates already known to be methylated. Whilst not all interacting proteins will be methyltransferase substrates it is clear methylation substrates of PMTs can be identified in the Y2H approach. SYNCRIP, EWSR1, FUS and SERBP1 are known to be methylated in RG-rich regions by PRMT1. SPBS2 interacting with

PRMT1 also contains a RG-rich region but is not know to be methylated thus presents a potential methylation candidate for PRMT1. We identified the lysine methylated WIZ protein interacting with the lysine methyltransferase SUV39H1 and the demethylase AOF2. WIZ and other proteins interacting with SUV39H1 are involved in transcriptional regulation. This study enables insights into SUV39H1 methylation function on non-histone proteins and identifies a set of potential substrates. The interaction network presents the first proteome-wide dataset of proteins interacting with enzymes associated with methylation and serves as a superior, well annotated resource to detect new methylation substrates and implicates protein methylation in several biological processes.

To determine sites of methylation we developed an *in vitro* methylation assay coupled to mass spectrometry. We validated our approach detecting previously unknown methylation sites on SPIN2B, DNAJA3, QKI, SAMD3, OFCC1, SYNCRIP and WDR42A. WDR42A and SPSB2 are involved in ubiquitination and proteasomal degradation. They function as substrate receptor and recruit proteins to be ubiquitinated. The methylation of WDR42A and SPSB2 could have regulatory function in recruitment of ubiquitination substrates. The presented interaction network and the generated models provide strong hypotheses for the molecular functions of methylation that build the basis for directed future investigation.

References

Abramovich, C., Yakobson, B., Chebath, J., and Revel, M. (1997). A protein-arginine methyltransferase binds to the intracytoplasmic domain of the IFNAR1 chain in the type I interferon receptor. The EMBO journal *16*, 260-266.

Adams, M.M., Wang, B., Xia, Z., Morales, J.C., Lu, X., Donehower, L.A., Bochar, D.A., Elledge, S.J., and Carpenter, P.B. (2005). 53BP1 oligomerization is independent of its methylation by PRMT1. Cell cycle *4*, 1854-1861.

Agger, K., Christensen, J., Cloos, P.A., and Helin, K. (2008). The emerging functions of histone demethylases. Current opinion in genetics & development *18*, 159-168.

An, W., Kim, J., and Roeder, R.G. (2004). Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. Cell *117*, 735-748.

Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., and Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature *443*, 590-593.

Araya, N., Hiraga, H., Kako, K., Arao, Y., Kato, S., and Fukamizu, A. (2005). Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. Biochemical and biophysical research communications *329*, 653-660.

Artzt, K., and Wu, J.I. (2010). STAR trek: An introduction to STAR family proteins and review of quaking (QKI). Advances in experimental medicine and biology *693*, 1-24.

Aso, T., Haque, D., Barstead, R.J., Conaway, R.C., and Conaway, J.W. (1996). The inducible elongin A elongation activation domain: structure, function and interaction with the elongin BC complex. The EMBO journal *15*, 5557-5566.

Bachand, F. (2007). Protein arginine methyltransferases: from unicellular eukaryotes to humans. Eukaryotic cell *6*, 889-898.

Bandyopadhyay, S., Chiang, C.Y., Srivastava, J., Gersten, M., White, S., Bell, R., Kurschner, C., Martin, C.H., Smoot, M., Sahasrabudhe, S.*, et al.* (2010). A human MAP kinase interactome. Nature methods *7*, 801-805.

Bannister, A.J., Schneider, R., and Kouzarides, T. (2002). Histone methylation: dynamic or static? Cell *109*, 801-806.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nature reviews Genetics *5*, 101-113.

Beck, H.C., Nielsen, E.C., Matthiesen, R., Jensen, L.H., Sehested, M., Finn, P., Grauslund, M., Hansen, A.M., and Jensen, O.N. (2006). Quantitative proteomic analysis of post-translational modifications of human histones. Molecular & cellular proteomics : MCP *5*, 1314-1325.

Bedford, D., McKeown, N., O'Farrell, A., and Howell, F. (2009). Alcohol levels in killed drivers and pedestrians on Irish roads 2003-2005: a national study. Irish medical journal *102*, 310, 312-314.

Bedford, M.T. (2007). Arginine methylation at a glance. Journal of cell science *120*, 4243-4246.

Bedford, M.T., and Clarke, S.G. (2009). Protein arginine methylation in mammals: who, what, and why. Molecular cell *33*, 1-13.

Bedford, M.T., Frankel, A., Yaffe, M.B., Clarke, S., Leder, P., and Richard, S. (2000). Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. The Journal of biological chemistry *275*, 16030-16036.

Bedford, M.T., and Richard, S. (2005). Arginine methylation an emerging regulator of protein function. Molecular cell *18*, 263-272.

Belyanskaya, L.L., Gehrig, P.M., and Gehring, H. (2001). Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. The Journal of biological chemistry *276*, 18681-18687.

Blanton, S., Srinivasan, A., and Rymond, B.C. (1992). PRP38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 small nuclear RNA levels. Molecular and cellular biology *12*, 3939-3947.

Boisvert, F.M., Chenard, C.A., and Richard, S. (2005a). Protein interfaces in signaling regulated by arginine methylation. Science's STKE : signal transduction knowledge environment *2005*, re2.

Boisvert, F.M., Cote, J., Boulanger, M.C., and Richard, S. (2003). A proteomic analysis of arginine-methylated protein complexes. Molecular & cellular proteomics : MCP *2*, 1319-1330.

Boisvert, F.M., Rhie, A., Richard, S., and Doherty, A.J. (2005b). The GAR motif of 53BP1 is arginine methylated by PRMT1 and is necessary for 53BP1 DNA binding activity. Cell cycle *4*, 1834-1841.

Boulanger, M.C., Liang, C., Russell, R.S., Lin, R., Bedford, M.T., Wainberg, M.A., and Richard, S. (2005). Methylation of Tat by PRMT6 regulates human immunodeficiency virus type 1 gene expression. Journal of virology *79*, 124-131.

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Luhrmann, R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. Rna *7*, 1531-1542.

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., and Luhrmann, R. (2000). The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. The Journal of biological chemistry *275*, 17122-17129.

Branscombe, T.L., Frankel, A., Lee, J.H., Cook, J.R., Yang, Z., Pestka, S., and Clarke, S. (2001). PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. The Journal of biological chemistry *276*, 32971-32976.

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J.M., Murray, R.R., Roncari, L., de Smet, A.S.*, et al.* (2009). An experimentally derived confidence score for binary protein-protein interactions. Nature methods *6*, 91-97.

Bussow, K., Scheich, C., Sievert, V., Harttig, U., Schultz, J., Simon, B., Bork, P., Lehrach, H., and Heinemann, U. (2005). Structural genomics of human proteins--target selection and generation of a public catalogue of expression clones. Microbial cell factories *4*, 21.

Cazalla, D., Newton, K., and Caceres, J.F. (2005). A novel SR-related protein is required for the second step of Pre-mRNA splicing. Molecular and cellular biology *25*, 2969-2980.

Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K. (2007). JMJD6 is a histone arginine demethylase. Science *318*, 444-447.

Chen, D., Ma, H., Hong, H., Koh, S.S., Huang, S.M., Schurter, B.T., Aswad, D.W., and Stallcup, M.R. (1999). Regulation of transcription by a protein methyltransferase. Science *284*, 2174-2177.

Cheng, D., Cote, J., Shaaban, S., and Bedford, M.T. (2007). The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. Molecular cell *25*, 71-83.

Cheng, H., Cenciarelli, C., Tao, M., Parks, W.P., and Cheng-Mayer, C. (2002). HTLV-1 Tax-associated hTid-1, a human DnaJ protein, is a repressor of Ikappa B kinase beta subunit. The Journal of biological chemistry *277*, 20605-20610.

Chevillard-Briet, M., Trouche, D., and Vandel, L. (2002). Control of CBP co-activating activity by arginine methylation. The EMBO journal *21*, 5457-5466.

Chi, T., Yan, Z., Xue, Y., and Wang, W. (2004). Purification and functional analysis of the mammalian SWI/SNF-family of chromatin-remodeling complexes. Methods in enzymology *377*, 299-316.

Chien, C.T., Bartel, P.L., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proceedings of the National Academy of Sciences of the United States of America *88*, 9578-9582.

Cho, H.S., Suzuki, T., Dohmae, N., Hayami, S., Unoki, M., Yoshimatsu, M., Toyokawa, G., Takawa, M., Chen, T., Kurash, J.K.*, et al.* (2011). Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells. Cancer research *71*, 655-660.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834-840.

Choudhary, C., and Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. Nature reviews Molecular cell biology *11*, 427-439.

Chua, K., and Reed, R. (1999). Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. Genes & development *13*, 841-850.

Chuikov, S., Kurash, J.K., Wilson, J.R., Xiao, B., Justin, N., Ivanov, G.S., McKinney, K., Tempst, P., Prives,

C., Gamblin, S.J., *et al.* (2004). Regulation of p53 activity through lysine methylation. Nature *432*, 353-360.

Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). *In vitro* and *in vivo* analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. The Journal of biological chemistry *280*, 5563-5570.

Cook, J.R., Lee, J.H., Yang, Z.H., Krause, C.D., Herth, N., Hoffmann, R., and Pestka, S. (2006). FBXO11/PRMT9, a new protein arginine methyltransferase, symmetrically dimethylates arginine residues. Biochemical and biophysical research communications *342*, 472-481.

Corsini, L., and Sattler, M. (2007). Tudor hooks up with DNA repair. Nature structural & molecular biology *14*, 98-99.

Cote, J., Boisvert, F.M., Boulanger, M.C., Bedford, M.T., and Richard, S. (2003). Sam68 RNA binding protein is an *in vivo* substrate for protein arginine N-methyltransferase 1. Molecular biology of the cell *14*, 274-287.

Covic, M., Hassa, P.O., Saccani, S., Buerki, C., Meier, N.I., Lombardi, C., Imhof, R., Bedford, M.T., Natoli, G., and Hottiger, M.O. (2005). Arginine methyltransferase CARM1 is a promoter-specific regulator of NF-kappaB-dependent gene expression. The EMBO journal *24*, 85-96.

Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J., *et al.* (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545-553.

Dacwag, C.S., Ohkawa, Y., Pal, S., Sif, S., and Imbalzano, A.N. (2007). The protein arginine methyltransferase Prmt5 is required for myogenesis because it facilitates ATP-dependent chromatin remodeling. Molecular and cellular biology *27*, 384-394.

Davies, S.J., Wise, C., Venkatesh, B., Mirza, G., Jefferson, A., Volpi, E.V., and Ragoussis, J. (2004). Mapping of three translocation breakpoints associated with orofacial clefting within 6p24 and identification of new transcripts within the region. Cytogenetic and genome research *105*, 47-53.

Dhayalan, A., Kudithipudi, S., Rathert, P., and Jeltsch, A. (2011). Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. Chemistry & biology *18*, 111-120.

Dickinson, L.A., Edgar, A.J., Ehley, J., and Gottesfeld, J.M. (2002). Cyclin L is an RS domain protein involved in pre-mRNA splicing. The Journal of biological chemistry *277*, 25465-25473.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome biology *6*, 227.

Duncan, P.I., Stojdl, D.F., Marius, R.M., Scheit, K.H., and Bell, J.C. (1998). The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing. Experimental cell research *241*, 300-308.

Ea, C.K., and Baltimore, D. (2009). Regulation of NF-kappaB activity through lysine monomethylation of p65. Proceedings of the National Academy of Sciences of the United States of America *106*, 18972-18977.

Ebersole, T.A., Chen, Q., Justice, M.J., and Artzt, K. (1996). The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. Nature genetics *12*, 260-265.

El-Andaloussi, N., Valovka, T., Toueille, M., Steinacher, R., Focke, F., Gehrig, P., Covic, M., Hassa, P.O., Schar, P., Hubscher, U., *et al.* (2006). Arginine methylation regulates DNA polymerase beta. Molecular cell *22*, 51-62.

El Messaoudi, S., Fabbrizio, E., Rodriguez, C., Chuchana, P., Fauquier, L., Cheng, D., Theillet, C., Vandel, L., Bedford, M.T., and Sardet, C. (2006). Coactivator-associated arginine methyltransferase 1 (CARM1) is a positive regulator of the Cyclin E1 gene. Proceedings of the National Academy of Sciences of the United States of America *103*, 13351-13356.

Esteve, P.O., Chang, Y., Samaranayake, M., Upadhyay, A.K., Horton, J.R., Feehery, G.R., Cheng, X., and Pradhan, S. (2011). A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nature structural & molecular biology *18*, 42-48.

Esteve, P.O., Chin, H.G., Benner, J., Feehery, G.R., Samaranayake, M., Horwitz, G.A., Jacobsen, S.E., and Pradhan, S. (2009). Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *106*, 5076-5081.

Eyckerman, S., Verhee, A., der Heyden, J.V., Lemmens, I., Ostade, X.V., Vandekerckhove, J., and Tavernier, J. (2001). Design and application of a cytokine-receptor-based interaction trap. Nature cell biology

3, 1114-1119.

Feng, Q., He, B., Jung, S.Y., Song, Y., Qin, J., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (2009). Biochemical control of CARM1 enzymatic activity by phosphorylation. The Journal of biological chemistry *284*, 36167-36174.

Fiedler, M., Sanchez-Barrena, M.J., Nekrasov, M., Mieszczanek, J., Rybin, V., Muller, J., Evans, P., and Bienz, M. (2008). Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. Molecular cell *30*, 507-518.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Figeys, D. (2008). Mapping the human protein interactome. Cell research *18*, 716-724.

Fletcher, B.S., Dragstedt, C., Notterpek, L., and Nolan, G.P. (2002). Functional cloning of SPIN-2, a nuclear anti-apoptotic protein with roles in cell cycle progression. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK *16*, 1507-1518.

Frankel, A., and Clarke, S. (2000). PRMT3 is a distinct member of the protein arginine N-methyltransferase family. Conferral of substrate specificity by a zinc-finger domain. The Journal of biological chemistry *275*, 32974-32982.

Frankel, A., Yadav, N., Lee, J., Branscombe, T.L., Clarke, S., and Bedford, M.T. (2002). The novel human protein arginine N-methyltransferase PRMT6 is a nuclear enzyme displaying unique substrate specificity. The Journal of biological chemistry *277*, 3537-3543.

Friesen, W.J., Wyce, A., Paushkin, S., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). A novel WD repeat protein component of the methylosome binds Sm proteins. The Journal of biological chemistry *277*, 8243-8247.

Frietze, S., Lupien, M., Silver, P.A., and Brown, M. (2008). CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1. Cancer research *68*, 301-306.

Fritsch, L., Robin, P., Mathieu, J.R., Souidi, M., Hinaux, H., Rougeulle, C., Harel-Bellan, A., Ameyar-Zazoua, M., and Ait-Si-Ali, S. (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Molecular cell *37*, 46-56.

Garrido, C., Schmitt, E., Cande, C., Vahsen, N., Parcellier, A., and Kroemer, G. (2003). HSP27 and HSP70: potentially oncogenic apoptosis inhibitors. Cell cycle *2*, 579-584.

Gary, J.D., and Clarke, S. (1998). RNA and protein interactions modulated by protein arginine methylation. Progress in nucleic acid research and molecular biology *61*, 65-131.

Gilbert, M., and Albala, J.S. (2002). Accelerating code to function: sizing up the protein production line. Current opinion in chemical biology *6*, 102-105.

Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nature reviews Molecular cell biology *8*, 645-654.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E.*, et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Gu, W., and Roeder, R.G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell *90*, 595-606.

Guo, Q., Bedford, M.T., and Fast, W. (2011). Discovery of peptidylarginine deiminase-4 substrates by protein array: antagonistic citrullination and methylation of human ribosomal protein S2. Molecular bioSystems *7*, 2286-2295.

Guo, Z., Zheng, L., Xu, H., Dai, H., Zhou, M., Pascua, M.R., Chen, Q.M., and Shen, B. (2010). Methylation of FEN1 suppresses nearby phosphorylation and facilitates PCNA binding. Nature chemical biology *6*, 766-773.

Hardy, R.J., Loushin, C.L., Friedrich, V.L., Jr., Chen, Q., Ebersole, T.A., Lazzarini, R.A., and Artzt, K. (1996). Neural cell type-specific expression of QKI proteins is altered in quakingviable mutant mice. The Journal of neuroscience : the official journal of the Society for Neuroscience *16*, 7941-7949.

Hassa, P.O., Covic, M., Bedford, M.T., and Hottiger, M.O. (2008). Protein arginine methyltransferase 1 coactivates NF-kappaB-dependent gene expression synergistically with CARM1 and PARP1. Journal of molecular biology *377*, 668-678.

Hastings, M.L., Allemand, E., Duelli, D.M., Myers, M.P., and Krainer, A.R. (2007). Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). PloS one *2*, e538.

He, N., Jahchan, N.S., Hong, E., Li, Q., Bayfield, M.A., Maraia, R.J., Luo, K., and Zhou, Q. (2008). A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. Molecular cell *29*, 588-599.

He, Y.J., McCall, C.M., Hu, J., Zeng, Y., and Xiong, Y. (2006). DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. Genes & development *20*, 2949-2954.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Luhrmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Molecular cell *45*, 567-580.

Herrmann, F., Pably, P., Eckerich, C., Bedford, M.T., and Fackelmayer, F.O. (2009). Human protein arginine methyltransferases *in vivo*--distinct properties of eight canonical members of the PRMT family. Journal of cell science *122*, 667-677.

Higa, L.A., Wu, M., Ye, T., Kobayashi, R., Sun, H., and Zhang, H. (2006). CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. Nature cell biology *8*, 1277-1283.

Hilton, D.J., Richardson, R.T., Alexander, W.S., Viney, E.M., Willson, T.A., Sprigg, N.S., Starr, R., Nicholson, S.E., Metcalf, D., and Nicola, N.A. (1998). Twenty proteins containing a C-terminal SOCS box form five structural classes. Proceedings of the National Academy of Sciences of the United States of America *95*, 114-119.

Hong, H., Kao, C., Jeng, M.H., Eble, J.N., Koch, M.O., Gardner, T.A., Zhang, S., Li, L., Pan, C.X., Hu, Z.*, et al.* (2004). Aberrant expression of CARM1, a transcriptional coactivator of androgen receptor, in the development of prostate carcinoma and androgen-independent status. Cancer *101*, 83-89.

Houtz, R.L., Stults, J.T., Mulligan, R.M., and Tolbert, N.E. (1989). Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase. Proceedings of the National Academy of Sciences of the United States of America *86*, 1855-1859.

Hsu, J.M., Chen, C.T., Chou, C.K., Kuo, H.P., Li, L.Y., Lin, C.Y., Lee, H.J., Wang, Y.N., Liu, M., Liao, H.W.*, et al.* (2011). Crosstalk between Arg 1175 methylation and Tyr 1173 phosphorylation negatively modulates EGFR-mediated ERK activation. Nature cell biology *13*, 174-181.

Huang, J., and Berger, S.L. (2008). The emerging field of dynamic lysine methylation of non-histone proteins. Current opinion in genetics & development *18*, 152-158.

Huang, J., Perez-Burgos, L., Placek, B.J., Sengupta, R., Richter, M., Dorsey, J.A., Kubicek, S., Opravil, S., Jenuwein, T., and Berger, S.L. (2006). Repression of p53 activity by Smyd2-mediated methylation. Nature *444*, 629-632.

Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M.T., Jenuwein, T.*, et al.* (2007). p53 is regulated by the lysine demethylase LSD1. Nature *449*, 105-108.

Hublitz, P., Albert, M., and Peters, A.H. (2009). Mechanisms of transcriptional repression by histone lysine methylation. The International journal of developmental biology *53*, 335-354.

Invernizzi, C.F., Xie, B., Frankel, F.A., Feldhammer, M., Roy, B.B., Richard, S., and Wainberg, M.A. (2007). Arginine methylation of the HIV-1 nucleocapsid protein results in its diminished function. Aids *21*, 795-805.

Invernizzi, C.F., Xie, B., Richard, S., and Wainberg, M.A. (2006). PRMT6 diminishes HIV-1 Rev binding to and export of viral RNA. Retrovirology *3*, 93.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences of the United States of America *98*, 4569-4574.

Iwai, K., Yamanaka, K., Kamura, T., Minato, N., Conaway, R.C., Conaway, J.W., Klausner, R.D., and Pause, A. (1999). Identification of the von Hippel-lindau tumor-suppressor protein as part of an active E3 ubiquitin ligase complex. Proceedings of the National Academy of Sciences of the United States of America *96*, 12436-12441.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jin, J., Arias, E.E., Chen, J., Harper, J.W., and Walter, J.C. (2006). A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1.

Molecular cell *23*, 709-721.

Kamura, T., Burian, D., Yan, Q., Schmidt, S.L., Lane, W.S., Querido, E., Branton, P.E., Shilatifard, A., Conaway, R.C., and Conaway, J.W. (2001). Muf1, a novel Elongin BC-interacting leucine-rich repeat protein that can assemble with Cul5 and Rbx1 to reconstitute a ubiquitin ligase. The Journal of biological chemistry *276*, 29748-29753.

Kamura, T., Sato, S., Haque, D., Liu, L., Kaelin, W.G., Jr., Conaway, R.C., and Conaway, J.W. (1998). The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. Genes & development *12*, 3872-3881.

Katsanis, N., Yaspo, M.L., and Fisher, E.M. (1997). Identification and mapping of a novel human gene, HRMT1L1, homologous to the rat protein arginine N-methyltransferase 1 (PRMT1) gene. Mammalian genome : official journal of the International Mammalian Genome Society *8*, 526-529.

Kibel, A., Iliopoulos, O., DeCaprio, J.A., and Kaelin, W.G., Jr. (1995). Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C. Science *269*, 1444-1446.

Kim, D., Lee, J., Cheng, D., Li, J., Carter, C., Richie, E., and Bedford, M.T. (2010). Enzymatic activity is required for the *in vivo* functions of CARM1. The Journal of biological chemistry *285*, 1147-1152.

Kim, J.D., Kako, K., Kakiuchi, M., Park, G.G., and Fukamizu, A. (2008). EWS is a substrate of type I protein arginine methyltransferase, PRMT8. International journal of molecular medicine *22*, 309-315.

Kim, S.C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L.*, et al.* (2006a). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Molecular cell *23*, 607-618.

Kim, S.M., Kee, H.J., Choe, N., Kim, J.Y., Kook, H., Kook, H., and Seo, S.B. (2007). The histone methyltransferase activity of WHISTLE is important for the induction of apoptosis and HDAC1-mediated transcriptional repression. Experimental cell research *313*, 975-983.

Kim, S.M., Kee, H.J., Eom, G.H., Choe, N.W., Kim, J.Y., Kim, Y.S., Kim, S.K., Kook, H., and Seo, S.B. (2006b). Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity. Biochemical and biophysical research communications *345*, 318-323.

Klose, R.J., and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. Nature reviews Molecular cell biology *8*, 307-318.

Kluck, R.M., Ellerby, L.M., Ellerby, H.M., Naiem, S., Yaffe, M.P., Margoliash, E., Bredesen, D., Mauk, A.G., Sherman, F., and Newmeyer, D.D. (2000). Determinants of cytochrome c pro-apoptotic activity. The role of lysine 72 trimethylation. The Journal of biological chemistry *275*, 16127-16133.

Klug, A., and Schwabe, J.W. (1995). Protein motifs 5. Zinc fingers. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology *9*, 597-604.

Koh, S.S., Li, H., Lee, Y.H., Widelitz, R.B., Chuong, C.M., and Stallcup, M.R. (2002). Synergistic coactivator function by coactivator-associated arginine methyltransferase (CARM) 1 and beta-catenin with two different classes of DNA-binding transcriptional activators. The Journal of biological chemistry *277*, 26031-26035.

Komyod, W., Bauer, U.M., Heinrich, P.C., Haan, S., and Behrmann, I. (2005). Are STATS arginine-methylated? The Journal of biological chemistry *280*, 21700-21705.

Kondo, T., Furuta, T., Mitsunaga, K., Ebersole, T.A., Shichiri, M., Wu, J., Artzt, K., Yamamura, K., and Abe, K. (1999). Genomic organization and expression analysis of the mouse qkI locus. Mammalian genome : official journal of the International Mammalian Genome Society *10*, 662-669.

Kouskouti, A., Scheer, E., Staub, A., Tora, L., and Talianidis, I. (2004). Gene-specific modulation of TAF10 function by SET9-mediated methylation. Molecular cell *14*, 175-182.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krause, C.D., Yang, Z.H., Kim, Y.S., Lee, J.H., Cook, J.R., and Pestka, S. (2007). Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. Pharmacology & therapeutics *113*, 50-87.

Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003). Structural classification of zinc fingers: survey and summary. Nucleic acids research *31*, 532-550.

Kruiswijk, T., Kunst, A., Planta, R.J., and Mager, W.H. (1978). Modification of yeast ribosomal proteins. Methylation. The Biochemical journal *175*, 221-225.

Kuang, Z., Lewis, R.S., Curtis, J.M., Zhan, Y., Saunders, B.M., Babon, J.J., Kolesnik, T.B., Low, A., Masters, S.L., Willson, T.A.*, et al.* (2010). The SPRY domain-containing SOCS box protein SPSB2 targets

iNOS for proteasomal degradation. The Journal of cell biology *190*, 129-141.

Kuhn, P., Chumanov, R., Wang, Y., Ge, Y., Burgess, R.R., and Xu, W. (2011). Automethylation of CARM1 allows coupling of transcription and mRNA splicing. Nucleic acids research *39*, 2717-2726.

Kunizaki, M., Hamamoto, R., Silva, F.P., Yamaguchi, K., Nagayasu, T., Shibuya, M., Nakamura, Y., and Furukawa, Y. (2007). The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3. Cancer research *67*, 10759-10765.

Kurash, J.K., Lei, H., Shen, Q., Marston, W.L., Granda, B.W., Fan, H., Wall, D., Li, E., and Gaudet, F. (2008). Methylation of p53 by Set7/9 mediates p53 acetylation and activity *in vivo*. Molecular cell *29*, 392-400.

Kurzik-Dumke, U., Debes, A., Kaymer, M., and Dienes, P. (1998). Mitochondrial localization and temporal expression of the Drosophila melanogaster DnaJ homologous tumor suppressor Tid50. Cell stress & chaperones *3*, 12-27.

Kwak, Y.T., Guo, J., Prajapati, S., Park, K.J., Surabhi, R.M., Miller, B., Gehrig, P., and Gaynor, R.B. (2003). Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. Molecular cell *11*, 1055-1066.

Kwiatkowski, T.J., Jr., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T.*, et al.* (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science *323*, 1205-1208.

Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. Current opinion in structural biology *11*, 39-46.

Lee, D.Y., Northrop, J.P., Kuo, M.H., and Stallcup, M.R. (2006). Histone H3 lysine 9 methyltransferase G9a is a transcriptional coactivator for nuclear receptors. The Journal of biological chemistry *281*, 8476-8485.

Lee, D.Y., Teyssier, C., Strahl, B.D., and Stallcup, M.R. (2005a). Role of protein methylation in regulation of transcription. Endocrine reviews *26*, 147-170.

Lee, J., and Bedford, M.T. (2002). PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. EMBO reports *3*, 268-273.

Lee, J., Cheng, D., and Bedford, M.T. (2004). Techniques in protein methylation. Methods in molecular biology *284*, 195-208.

Lee, J., Sayegh, J., Daniel, J., Clarke, S., and Bedford, M.T. (2005b). PRMT8, a new membrane-bound tissue-specific member of the protein arginine methyltransferase family. The Journal of biological chemistry *280*, 32890-32896.

Lee, J., and Zhou, P. (2007). DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. Molecular cell *26*, 775-780.

Lee, J.H., Cook, J.R., Yang, Z.H., Mirochnitchenko, O., Gunderson, S.I., Felix, A.M., Herth, N., Hoffmann, R., and Pestka, S. (2005c). PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. The Journal of biological chemistry *280*, 3656-3664.

Lee, Y.H., Coonrod, S.A., Kraus, W.L., Jelinek, M.A., and Stallcup, M.R. (2005d). Regulation of coactivator complex assembly and function by protein arginine methylation and demethylimination. Proceedings of the National Academy of Sciences of the United States of America *102*, 3611-3616.

Lee, Y.H., and Stallcup, M.R. (2009). Minireview: protein arginine methylation of nonhistone proteins in transcriptional regulation. Molecular endocrinology *23*, 425-433.

Leon, O., and Roth, M. (2000). Zinc fingers: DNA binding and protein-protein interactions. Biological research *33*, 21-30.

Levy, D., Kuo, A.J., Chang, Y., Schaefer, U., Kitson, C., Cheung, P., Espejo, A., Zee, B.M., Liu, C.L., Tangsombatvisit, S.*, et al.* (2011). Lysine methylation of the NF-kappaB subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-kappaB signaling. Nature immunology *12*, 29-36.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T.*, et al.* (2004). A map of the interactome network of the metazoan C. elegans. Science *303*, 540-543.

Li, T., and Kelly, W.G. (2011). A role for Set1/MLL-related components in epigenetic regulation of the Caenorhabditis elegans germ line. PLoS genetics *7*, e1001349.

Li, Z.Z., Kondo, T., Murata, T., Ebersole, T.A., Nishi, T., Tada, K., Ushio, Y., Yamamura, K., and Abe, K. (2002). Expression of Hqk encoding a KH RNA binding protein is altered in human glioma. Japanese journal

of cancer research : Gann *93*, 167-177.

Lin, W.J., Gary, J.D., Yang, M.C., Clarke, S., and Herschman, H.R. (1996). The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. The Journal of biological chemistry *271*, 15034-15044.

Liu, H.Y., MacDonald, J.I., Hryciw, T., Li, C., and Meakin, S.O. (2005). Human tumorous imaginal disc 1 (TID1) associates with Trk receptor tyrosine kinases and regulates neurite outgrowth in nnr5-TrkA cells. The Journal of biological chemistry *280*, 19461-19471.

Liu, Q., and Dreyfuss, G. (1995). In vivo and *in vitro* arginine methylation of RNA-binding proteins. Molecular and cellular biology *15*, 2800-2808.

Loh, Y.H., Zhang, W., Chen, X., George, J., and Ng, H.H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes & development *21*, 2545-2557.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nature reviews Molecular cell biology *6*, 838-849.

Martin, J.L., and McMillan, F.M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Current opinion in structural biology *12*, 783-793.

Masters, S.L., Yao, S., Willson, T.A., Zhang, J.G., Palmer, K.R., Smith, B.J., Babon, J.J., Nicola, N.A., Norton, R.S., and Nicholson, S.E. (2006). The SPRY domain of SSB-2 adopts a novel fold that presents conserved Par-4-binding residues. Nature structural & molecular biology *13*, 77-84.

McBride, A.E., and Silver, P.A. (2001). State of the arg: protein methylation at arginine comes of age. Cell *106*, 5-8.

Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C., and Fischer, U. (2001). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Current biology : CB *11*, 1990-1994.

Meister, G., and Fischer, U. (2002). Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. The EMBO journal *21*, 5853-5863.

Meyer, R., Wolf, S.S., and Obendorf, M. (2007). PRMT2, a member of the protein arginine methyltransferase family, is a coactivator of the androgen receptor. The Journal of steroid biochemistry and molecular biology *107*, 1-14.

Miranda, T.B., Miranda, M., Frankel, A., and Clarke, S. (2004). PRMT7 is a member of the protein arginine methyltransferase family with a distinct substrate specificity. The Journal of biological chemistry *279*, 22902-22907.

Miranda, T.B., Webb, K.J., Edberg, D.D., Reeves, R., and Clarke, S. (2005). Protein arginine methyltransferase 6 specifically methylates the nonhistone chromatin protein HMGA1a. Biochemical and biophysical research communications *336*, 831-835.

Najbauer, J., and Aswad, D.W. (1990). Diversity of methyl acceptor proteins in rat pheochromocytoma (PC12) cells revealed after treatment with adenosine dialdehyde. The Journal of biological chemistry *265*, 12717-12721.

Najbauer, J., Johnson, B.A., Young, A.L., and Aswad, D.W. (1993). Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins. The Journal of biological chemistry *268*, 10501-10509.

Nicholson, S.E., and Hilton, D.J. (1998). The SOCS proteins: a new family of negative regulators of signal transduction. Journal of leukocyte biology *63*, 665-668.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes & development *16*, 479-489.

O'Carroll, D., Scherthan, H., Peters, A.H., Opravil, S., Haynes, A.R., Laible, G., Rea, S., Schmid, M., Lebersorger, A., Jerratsch, M.*, et al.* (2000). Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. Molecular and cellular biology *20*, 9423-9433.

Ong, S.E., Mittler, G., and Mann, M. (2004). Identifying and quantifying *in vivo* methylation sites by heavy methyl SILAC. Nature methods *1*, 119-126.

Pahlich, S., Zakaryan, R.P., and Gehring, H. (2006). Protein arginine methylation: Cellular functions and methods of analysis. Biochimica et biophysica acta *1764*, 1890-1903.

Pahlich, S., Zakaryan, R.P., and Gehring, H. (2008). Identification of proteins interacting with protein arginine methyltransferase 8: the Ewing sarcoma (EWS) protein binds independent of its methylation state. Proteins *72*, 1125-1137.

Passos, D.O., Bressan, G.C., Nery, F.C., and Kobarg, J. (2006a). Ki-1/57 interacts with PRMT1 and is a substrate for arginine methylation. The FEBS journal *273*, 3946-3961.

Passos, D.O., Quaresma, A.J., and Kobarg, J. (2006b). The methylation of the C-terminal region of hnRNPQ (NSAP1) is important for its nuclear localization. Biochemical and biophysical research communications *346*, 517-525.

Pawlak, M.R., Banik-Maiti, S., Pietenpol, J.A., and Ruley, H.E. (2002). Protein arginine methyltransferase I: substrate specificity and role in hnRNP assembly. Journal of cellular biochemistry *87*, 394-407.

Pawlak, M.R., Scherer, C.A., Chen, J., Roshon, M.J., and Ruley, H.E. (2000). Arginine N-methyltransferase 1 is required for early postimplantation mouse development, but cells deficient in the enzyme are viable. Molecular and cellular biology *20*, 4859-4869.

Peng, H., Begg, G.E., Harper, S.L., Friedman, J.R., Speicher, D.W., and Rauscher, F.J., 3rd (2000). Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. The Journal of biological chemistry *275*, 18000-18010.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589.

Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. Nature reviews Molecular cell biology *6*, 9-20.

Petrossian, T.C., and Clarke, S.G. (2011). Uncovering the human methyltransferasome. Molecular & cellular proteomics : MCP *10*, M110 000976.

Piessevaux, J., Lavens, D., Peelman, F., and Tavernier, J. (2008). The many faces of the SOCS box. Cytokine & growth factor reviews *19*, 371-381.

Polevoda, B., Martzen, M.R., Das, B., Phizicky, E.M., and Sherman, F. (2000). Cytochrome c methyltransferase, Ctm1p, of yeast. The Journal of biological chemistry *275*, 20508-20513.

Pollack, B.P., Kotenko, S.V., He, W., Izotova, L.S., Barnoski, B.L., and Pestka, S. (1999). The human homologue of the yeast proteins Skb1 and Hsl7p interacts with Jak kinases and contains protein methyltransferase activity. The Journal of biological chemistry *274*, 31531-31542.

Porras-Yakushi, T.R., Whitelegge, J.P., Miranda, T.B., and Clarke, S. (2005). A novel SET domain methyltransferase modifies ribosomal protein Rpl23ab in yeast. The Journal of biological chemistry *280*, 34590-34598.

Qi, C., Chang, J., Zhu, Y., Yeldandi, A.V., Rao, S.M., and Zhu, Y.J. (2002). Identification of protein arginine methyltransferase 2 as a coactivator for estrogen receptor alpha. The Journal of biological chemistry *277*, 28624-28630.

Rajagopala, S.V., Hughes, K.T., and Uetz, P. (2009). Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins. Proteomics *9*, 5296-5302.

Rappsilber, J., Friesen, W.J., Paushkin, S., Dreyfuss, G., and Mann, M. (2003). Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. Analytical chemistry *75*, 3107-3114.

Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008). Protein lysine methyltransferase G9a acts on non-histone targets. Nature chemical biology *4*, 344-346.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R.*, et al.* (2003). C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nature genetics *34*, 35-41.

Robin, P., Fritsch, L., Philipot, O., Svinarchuk, F., and Ait-Si-Ali, S. (2007). Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a. Genome biology *8*, R270.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D.,

Dreze, M., Ayivi-Guedehoussou, N., *et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Rudner, D.Z., Kanaar, R., Breger, K.S., and Rio, D.C. (1998). Interaction between subunits of heterodimeric splicing factor U2AF is essential *in vivo*. Molecular and cellular biology *18*, 1765-1773.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell *25*, 15-30.

Sampath, S.C., Marazzi, I., Yap, K.L., Sampath, S.C., Krutchinsky, A.N., Mecklenbrauker, I., Viale, A., Rudensky, E., Zhou, M.M., Chait, B.T., *et al.* (2007). Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. Molecular cell *27*, 596-608.

Sanderson, C.M. (2009). The Cartographers toolbox: building bigger and better human protein interaction networks. Briefings in functional genomics & proteomics *8*, 1-11.

Sarkar, S., Pollack, B.P., Lin, K.T., Kotenko, S.V., Cook, J.R., Lewis, A., and Pestka, S. (2001). hTid-1, a human DnaJ protein, modulates the interferon signaling pathway. The Journal of biological chemistry *276*, 49034-49042.

Sayegh, J., Webb, K., Cheng, D., Bedford, M.T., and Clarke, S.G. (2007). Regulation of protein arginine methyltransferase 8 (PRMT8) activity by its N-terminal domain. The Journal of biological chemistry *282*, 36444-36453.

Schotta, G., Ebert, A., and Reuter, G. (2003). SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. Genetica *117*, 149-158.

Schubert, H.L., Blumenthal, R.M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. Trends in biochemical sciences *28*, 329-335.

Schurter, B.T., Koh, S.S., Chen, D., Bunick, G.J., Harp, J.M., Hanson, B.L., Henschen-Edman, A., Mackay, D.R., Stallcup, M.R., and Aswad, D.W. (2001). Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. Biochemistry *40*, 5747-5756.

Schwartz, A.S., Yu, J., Gardenour, K.R., Finley, R.L., Jr., and Ideker, T. (2009). Cost-effective strategies for completing the interactome. Nature methods *6*, 55-61.

Scott, H.S., Antonarakis, S.E., Lalioti, M.D., Rossier, C., Silver, P.A., and Henry, M.F. (1998). Identification and characterization of two putative human arginine methyltransferases (HRMT1L1 and HRMT1L2). Genomics *48*, 330-340.

Shi, X., Kachirskaia, I., Yamaguchi, H., West, L.E., Wen, H., Wang, E.W., Dutta, S., Appella, E., and Gozani, O. (2007). Modulation of p53 function by SET8-mediated methylation at lysine 382. Molecular cell *27*, 636-646.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R., and Kielkopf, C.L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. Molecular cell *23*, 49-59.

Simonis, N., Rual, J.F., Carvunis, A.R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Sahalie, J.M., Venkatesan, K., Gebreab, F., *et al.* (2009). Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. Nature methods *6*, 47-54.

Sims, R.J., 3rd, Rojas, L.A., Beck, D., Bonasio, R., Schuller, R., Drury, W.J., 3rd, Eick, D., and Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science *332*, 99-103.

Sitaramayya, A., Wright, L.S., and Siegel, F.L. (1980). Enzymatic methylation of calmodulin in rat brain cytosol. The Journal of biological chemistry *255*, 8894-8900.

Stallcup, M.R. (2001). Role of protein methylation in chromatin remodeling and transcriptional regulation. Oncogene *20*, 3014-3020.

Stebbins, C.E., Kaelin, W.G., Jr., and Pavletich, N.P. (1999). Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science *284*, 455-461.

Stelzl, U., and Wanker, E.E. (2006). The value of high quality protein-protein interaction networks for systems biology. Current opinion in chemical biology *10*, 551-558.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., *et al.* (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell *122*, 957-968.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiology and molecular biology reviews : MMBR *64*, 435-459.

Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D.R., Collins, R.E., Sharma, D., Peng, J., Cheng, X., and Vertino, P.M. (2008). Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. Molecular cell *30*, 336-347.

Suter, B., Kittanakom, S., and Stagljar, I. (2008). Two-hybrid technologies in proteomics research. Current opinion in biotechnology *19*, 316-323.

Swiercz, R., Cheng, D., Kim, D., and Bedford, M.T. (2007). Ribosomal protein rpS2 is hypomethylated in PRMT3-deficient mice. The Journal of biological chemistry *282*, 16917-16923.

Swiercz, R., Person, M.D., and Bedford, M.T. (2005). Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). The Biochemical journal *386*, 85-91.

Syken, J., De-Medina, T., and Munger, K. (1999). TID1, a human homolog of the Drosophila tumor suppressor l(2)tid, encodes two mitochondrial modulators of apoptosis with opposing functions. Proceedings of the National Academy of Sciences of the United States of America *96*, 8499-8504.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development *19*, 815-826.

Tan, X., Rotllant, J., Li, H., De Deyne, P., and Du, S.J. (2006). SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos. Proceedings of the National Academy of Sciences of the United States of America *103*, 2713-2718.

Tang, J., Frankel, A., Cook, R.J., Kim, S., Paik, W.K., Williams, K.R., Clarke, S., and Herschman, H.R. (2000a). PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. The Journal of biological chemistry *275*, 7723-7730.

Tang, J., Gary, J.D., Clarke, S., and Herschman, H.R. (1998). PRMT 3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. The Journal of biological chemistry *273*, 16935-16945.

Tang, J., Kao, P.N., and Herschman, H.R. (2000b). Protein-arginine methyltransferase I, the predominant protein-arginine methyltransferase in cells, interacts with and is regulated by interleukin enhancer-binding factor 3. The Journal of biological chemistry *275*, 19866-19876.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature structural & molecular biology *14*, 1025-1040.

Tradewell, M.L., Yu, Z., Tibshirani, M., Boulanger, M.C., Durham, H.D., and Richard, S. (2012). Arginine methylation by PRMT1 regulates nuclear-cytoplasmic localization and toxicity of FUS/TLS harbouring ALS-linked mutations. Human molecular genetics *21*, 136-149.

Trentin, G.A., Yin, X., Tahir, S., Lhotak, S., Farhang-Fallah, J., Li, Y., and Rozakis-Adcock, M. (2001). A mouse homologue of the Drosophila tumor suppressor l(2)tid gene defines a novel Ras GTPase-activating protein (RasGAP)-binding protein. The Journal of biological chemistry *276*, 13087-13095.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Ueda, J., Tachibana, M., Ikura, T., and Shinkai, Y. (2006). Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. The Journal of biological chemistry *281*, 20120-20128.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.*, et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. Genome biology *4*, 231.

Van Langenhove, T., van der Zee, J., Sleegers, K., Engelborghs, S., Vandenberghe, R., Gijselinck, I., Van den Broeck, M., Mattheijssens, M., Peeters, K., De Deyn, P.P.*, et al.* (2010). Genetic contribution of FUS to frontotemporal lobar degeneration. Neurology *74*, 366-371.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P.*, et al.* (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science *323*, 1208-1211.

Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., *et al.* (2009). An empirical framework for binary interactome mapping. Nature methods *6*, 83-90.

Vernet, C., and Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of RNA. Trends in genetics : TIG *13*, 479-484.

Vinayagam, A., Stelzl, U., Foulle, R., Plassmann, S., Zenkner, M., Timm, J., Assmus, H.E., Andrade-Navarro, M.A., and Wanker, E.E. (2011). A directed protein interaction network for investigating intracellular signal transduction. Science signaling *4*, rs8.

Vinayagam, A., Stelzl, U., and Wanker, E.E. (2010). Repeat two-hybrid screening detects transient protein-protein interactions. Theor Chem Acc *125*, 613-619.

Wang, C., and Meier, U.T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. The EMBO journal *23*, 1857-1867.

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Molecular cell *8*, 1207-1217.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., *et al.* (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nature genetics *41*, 125-129.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.H., Perlin, J.R., Leonelli, L., Sonbuchner, L.S., McDonald, C.H., Cook, R.G., Dou, Y., *et al.* (2004). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science *306*, 279-283.

Webb, K.J., Laganowsky, A., Whitelegge, J.P., and Clarke, S.G. (2008). Identification of two SET domain proteins required for methylation of lysine residues in yeast ribosomal protein Rpl42ab. The Journal of biological chemistry *283*, 35561-35568.

Webby, C.J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M.L., Schmitz, C., Butler, D.S., Yates, J.R., 3rd, *et al.* (2009). Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. Science *325*, 90-93.

Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., *et al.* (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell *125*, 467-481.

Williams, A.J., Blacklow, S.C., and Collins, T. (1999). The zinc finger-associated SCAN box is a conserved oligomerization domain. Molecular and cellular biology *19*, 8526-8535.

Williams, A.J., Khachigian, L.M., Shows, T., and Collins, T. (1995). Isolation and characterization of a novel zinc-finger protein with transcription repressor activity. The Journal of biological chemistry *270*, 22143-22152.

Worseck, J.M., Grossmann, A., Weimann, M., Hegele, A., and Stelzl, U. (2012). A stringent yeast two-hybrid matrix screening approach for protein-protein interaction discovery. Methods in molecular biology *812*, 63-87.

Wu, J., Zhou, L., Tonissen, K., Tee, R., and Artzt, K. (1999). The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. The Journal of biological chemistry *274*, 29202-29210.

Xie, B., Invernizzi, C.F., Richard, S., and Wainberg, M.A. (2007). Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat Interactions with both cyclin T1 and the Tat transactivation region. Journal of virology *81*, 4226-4234.

Xu, W., Chen, H., Du, K., Asahara, H., Tini, M., Emerson, B.M., Montminy, M., and Evans, R.M. (2001). A transcriptional switch mediated by cofactor methylation. Science *294*, 2507-2511.

Yadav, N., Lee, J., Kim, J., Shen, J., Hu, M.C., Aldaz, C.M., and Bedford, M.T. (2003). Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. Proceedings of the National Academy of Sciences of the United States of America *100*, 6464-6468.

Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell *125*, 483-495.

Yang, X.D., Huang, B., Li, M., Lamb, A., Kelleher, N.L., and Chen, L.F. (2009). Negative regulation of NF-

kappaB action by Set9-mediated lysine methylation of the RelA subunit. The EMBO journal *28*, 1055-1066.

Yang, X.J., and Seto, E. (2008). Lysine acetylation: codified crosstalk with other posttranslational modifications. Molecular cell *31*, 449-461.

Yoshimoto, T., Boehm, M., Olive, M., Crook, M.F., San, H., Langenickel, T., and Nabel, E.G. (2006). The arginine methyltransferase PRMT2 binds RB and regulates E2F function. Experimental cell research *312*, 2040-2053.

Yoshimura, A., Naka, T., and Kubo, M. (2007). SOCS proteins, cytokine signalling and immune regulation. Nature reviews Immunology *7*, 454-465.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N.*, et al.* (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104-110.

Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X.*, et al.* (2011). Next-generation sequencing to generate interactome datasets. Nature methods *8*, 478-480.

Zhang, J.G., Farley, A., Nicholson, S.E., Willson, T.A., Zugaro, L.M., Simpson, R.J., Moritz, R.L., Cary, D., Richardson, R., Hausmann, G.*, et al.* (1999). The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. Proceedings of the National Academy of Sciences of the United States of America *96*, 2071-2076.

Zhang, J.G., Metcalf, D., Rakar, S., Asimakis, M., Greenhalgh, C.J., Willson, T.A., Starr, R., Nicholson, S.E., Carter, W., Alexander, W.S.*, et al.* (2001). The SOCS box of suppressor of cytokine signaling-1 is important for inhibition of cytokine action *in vivo*. Proceedings of the National Academy of Sciences of the United States of America *98*, 13261-13265.

Zhang, X., and Cheng, X. (2003). Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. Structure *11*, 509-520.

Zhang, Y., Lu, Z., Ku, L., Chen, Y., Wang, H., and Feng, Y. (2003). Tyrosine phosphorylation of QKI mediates developmental signals to regulate mRNA metabolism. The EMBO journal *22*, 1801-1810.

Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes & development *10*, 1356-1368.

Appendix

Table 6: PRMTs, PKMTs and PDeMs clones used in the different screening approaches as baits

GeneID	Symbol	ProtID	matrix	seq1	seq2	seq3
23028	AOF2	CCSB_10726	matrix	seq1	seq2	seq3
23028	AOF2	IOH26899	matrix	seq1	seq2	seq3
10498	CARM1	GC-Y3476-CF			seq2	seq3
10498	CARM1	GC-Y3476			seq2	seq3
79813	EHMT1	IOH26527	matrix	seq1	seq2	seq3
55818	JMJD1A	CCSB_1715	matrix	seq1	seq2	seq3
51780	JMJD1B	CCSB_3376	matrix	seq1	seq2	seq3
221037	JMJD1C	CCSB_5189	matrix	seq1	seq2	seq3
9682	JMJD2A	FLJ82489AAAF			seq2	seq3
23030	JMJD2B	CCSB_14302	matrix	seq1	seq2	seq3
65094	JMJD4	CCSB_52958			seq2	seq3
79831	JMJD5	CCSB_8354	matrix	seq1	seq2	seq3
23210	JMJD6	CCSB_12305	matrix	seq1	seq2	seq3
8681	JMJD7	CCSB_12694	matrix	seq1	seq2	seq3
3276	PRMT1	RZPDo834A082D	matrix	seq1	seq2	seq3
3276	PRMT1	ORF_SEQ_2A08	matrix	seq1	seq2	seq3
3276	PRMT1	CCSB_56119	matrix	seq1	seq2	seq3
3275	PRMT2	CCSB_3556	matrix	seq1	seq2	seq3
3275	PRMT2	RZPDo839A03141	matrix	seq1	seq2	seq3
3275	PRMT2	RZPDo834E0531D	matrix	seq1	seq2	seq3
10196	PRMT3	MWM1016	matrix	seq1	seq2	seq3
10196	PRMT3	MWM1026	matrix	seq1	seq2	seq3
10419	PRMT5	RZPDo839F0381			seq2	seq3
10419	PRMT5	AKHAMPS07n09			seq2	seq3
10419	PRMT5	RZPDo834G0114D			seq2	
55170	PRMT6	MWM1025	matrix	seq1	seq2	seq3
55170	PRMT6	CCSB_55527	matrix	seq1	seq2	seq3
54496	PRMT7	CCSB_45	matrix	seq1	seq2	seq3
56341	PRMT8	CCSB_1137	matrix	seq1	seq2	seq3
56341	PRMT8	RZPDo839E12147	matrix	seq1	seq2	seq3
56341	PRMT8	RZPDo839E12148	matrix	seq1	seq2	seq3
84193	SETD3	CCSB_11000	matrix	seq1	seq2	seq3
84193	SETD3	CCSB_8806	matrix	seq1	seq2	seq3
54093	SETD4	CCSB_4401			seq2	seq3
55209	SETD5	CCSB_524	matrix	seq1	seq2	seq3
79918	SETD6	CCSB_1153	matrix	seq1	seq2	seq3
80854	SETD7	MWM1015			seq2	seq3
387893	SETD8	CCSB_10598	matrix	seq1	seq2	seq3
83852	SETDB2	CCSB_10639	matrix	seq1	seq2	seq3
6419	SETMAR	CCSB_6228			seq2	seq3
6419	SETMAR	ORF_SEQ_3H12			seq2	seq3
150572	SMYD1	CCSB_54271			seq2	seq3
56950	SMYD2	CCSB_52672	matrix	seq1	seq2	
64754	SMYD3	CCSB_495			seq2	seq3
64754	SMYD3	RZPDo839C07149			seq2	seq3
64754	SMYD3	CCSB_13232			seq2	seq3
114826	SMYD4	CCSB_2337			seq2	seq3
6839	SUV39H1	RZPDo839A0973	matrix	seq1	seq2	seq3
6839	SUV39H1	RZPDo839A0963	matrix	seq1	seq2	seq3
6839	SUV39H1	CCSB_4830	matrix	seq1	seq2	seq3
79723	SUV39H2	ORF_SEQ_14C02	matrix	seq1	seq2	seq3
79723	SUV39H2	CCSB_4919	matrix	seq1	seq2	seq3
79723	SUV39H2	RZPDo839F1264	matrix	seq1	seq2	seq3
51111	SUV420H1	CCSB_55735			seq2	seq3
51111	SUV420H1	CCSB_9187			seq2	seq3
54904	WHSC1L1	CCSB_9120			seq2	seq3